









178  
9.11.76















# JOURNAL OF CONSULTING PSYCHOLOGY

Edited By  
Edward S. Bordin  
*University of Michigan*

Associate Editor  
Boyd McCandless  
*Indiana University*

Advisory Editors

John W. Atkinson  
*University of Michigan*

Frank X. Barron  
*University of California, Berkeley*

John M. Butler  
*University of Chicago*

Bertram D. Cohen  
*Rutgers University*

Allen T. Dittmann  
*National Institute of Mental Health*

Seymour Fisher  
*State University of New York*

Donald W. Fiske  
*University of Chicago*

Harrison G. Gough  
*University of California, Berkeley*

William L. Hays  
*University of Michigan*

Jerome Kagan  
*Fels Research Institute*

Leon H. Levy  
*Indiana University*

Paul E. Meehl  
*University of Minnesota*

Daniel R. Miller  
*University of Michigan*

Paul H. Mussen  
*University of California, Berkeley*

Alan K. Rosenwald  
*Illinois State Psychiatric Institute*

Saul Rosenzweig  
*Washington University*

Julian B. Rotter  
*Ohio State University*

Nevitt Sanford  
*Stanford University, Stanford*

Edwin S. Shneidman  
*Veterans Administration, Los Angeles, California*

Read D. Tuddenham  
*University of California, Berkeley*

George S. Welsh  
*University of North Carolina*

## VOLUME 27

## 1963

Helen Orr  
Managing Editor

Elizabeth S. Reed  
Advertising Manager

Rita Kopanic  
Editorial Assistant

Published Bimonthly By

The American Psychological Association, Inc.  
1333 Sixteenth St., N.W.  
Washington, D. C. 20036

Printed at  
Lancaster, Pennsylvania 17604

Copyright © 1963 by the American Psychological Association, Inc.





# Contents of Volume 27

<i>Adams, Henry E., and Kirby, Albert C.</i> Manifest Anxiety, Social Desirability, or Response Set - - - - -	59
<i>Agnew, Neil McK.</i> See <i>Pyke, Sandra.</i>	
<i>Albee, George W., Lane, Ellen A., Corcoran, Clare, and Werneke, Ann.</i> Childhood and Inter-current Intellectual Performance of Adult Schizophrenics - - - - -	364
<i>Allison, Roger B., Jr.</i> A Two-Dimensional Semantic Differential - - - - -	18
<i>Ambler, Rosalie K.</i> See <i>Willingham, Warren W.</i>	
<i>Ammons, Helen.</i> See <i>Johannsen, Walter J.</i>	
<i>Amon, Albert H.</i> See <i>Meer, Bernard.</i>	
<i>Anthony, Aristotle A.</i> The Relationship between Neuroticism, Stress, and Qualitative Porteus Maze Performance - - - - -	513
<i>Barrett, Beatrice H.</i> See <i>Zuckerman, Marvin.</i>	
<i>Bass, Bernard M.</i> See <i>Kanfer, Frederick H.</i>	
<i>Be'doch, Michael.</i> Applicability of the Norms of the Wechsler Intelligence Scale for Children to Five-Year-Olds - - - - -	263
<i>Bentler, P. M.</i> Interpersonal Orientation in Relation to Hypnotic Susceptibility - - -	426
<i>Bernstein, Lewis, and Dana, Richard H.</i> Effect of Order of Presentation of TAT Cards -	533
<i>Bloom, Bernard L.</i> Definitional Aspects of the Crisis Concept - - - - -	498
<i>Blumberg, Stanley.</i> See <i>Korman, Maurice.</i>	
<i>Boe, Erling E., and Kogan, William.</i> Social Desirability Response Set in the Individual -	369
<i>Bragg, Robert A.</i> See <i>Scott, Thomas R.</i>	
<i>Brilliant, Patricia J., and Gynther, Malcolm D.</i> Relationships between Performance on Three Tests for Organicity and Selected Patient Variables - - - - -	474
<i>Bryne, Donn, Golightly, Carole, and Capaldi, E. J.</i> Construction and Validation of the Food Attitude Scale - - - - -	215
<i>Caine, T. M., and Hawkins, L. G.</i> Questionnaire Measure of the Hysteroid/Obsessoid Component of Personality: The HOQ - - - - -	206
<i>Callahan, Daniel M.</i> See <i>McNair, Douglas M.</i>	
<i>Canter, Francis M.</i> Simulation on the California Psychological Inventory and the Adjustment of the Simulator - - - - -	253
<i>Capaldi, E. F.</i> See <i>Byrne, Donn.</i>	
<i>Caputo, Daniel V., Edmonston, William E., Jr., L'Abate, Luciano, and Rondberg, Samuel R.</i> Type of Brain Damage and Intellectual Functioning in Children - - - - -	184
<i>Cartwright, Rosalind Dymond, and Lerner, Barbara.</i> Empathy, Need to Change and Improvement with Psychotherapy - - - - -	138
<i>Centers, Louise, and Centers, Richard.</i> Body Cathexes of Parents of Normal and Malformed Children for Progeny and Self - - - - -	319
<i>Centers, Richard.</i> See <i>Centers, Louise.</i>	
<i>Chapman, Loren J.</i> The Problem of Selecting Drug-Free Schizophrenics for Research - -	540
<i>Chipman, Abram.</i> See <i>Siller, Jerome.</i>	
<i>Clore, Gerald L., Jr.</i> Kent E-G-Y: Differential Scoring and Correlation with the WAIS -	372
<i>Clyde, Dean H.</i> See <i>Raskin, Allen.</i>	
<i>Cohen, Benjamin.</i> See <i>Dibner, Andrew S.</i>	
<i>Corcoran, Clare.</i> See <i>Albee, George W.</i>	
<i>Costa, Louis D., Vaughan, Herbert G., Jr., Levita, Eric, and Farber, Norman.</i> Purdue Peg-board as a Predictor of the Presence and Laterality of Cerebral Lesions - - - - -	133
<i>Crookes, T. G., and Hull, S. J.</i> Scores of Psychotic Patients on the Maudsley Personality Inventory - - - - -	243

<i>Crowne, Douglas P.</i> See <i>Strickland, Bonnie R.</i>	
<i>Curtis, Floyd J.</i> See <i>Medinnus, Gene R.</i>	
<i>Dahlke, Arnold E., and Dana, Richard H.</i> Intraindividual Verbal-Numerical Discrepancies and Personality - - - - -	182
<i>Dana, Richard H.</i> See <i>Bernstein, Lewis.</i>	
<i>Dana, Richard H.</i> See <i>Dahlke, Arnold E.</i>	
<i>Deering, Gayle.</i> Affective Stimuli and Disturbance of Thought Processes - - - - -	338
<i>Dempsey, Paul.</i> The Dimensionality of the MMPI Clinical Scales among Normal Subjects	492
<i>den Breeijen, Arie.</i> See <i>Zuckerman, Marvin.</i>	
<i>Dibner, Andrew S., Palmer, Robert D., Cohen, Benjamin, and Gofstein, Arnold G.</i> The Use of an Open-Ended Group in the Intake Procedure of a Mental Hygiene Unit - - - -	83
<i>Domino, George.</i> See <i>Gough, Harrison G.</i>	
<i>Edmonston, William E., Jr.</i> See <i>Caputo, Daniel V.</i>	
<i>Eells, Janet.</i> See <i>Feifel, Herman.</i>	
<i>Eliseo, Thomas S.</i> Overinclusive Thinking in Process and Reactive Schizophrenics - - -	447
<i>Elvekrog, Maurice O., and Vestre, Norris D.</i> The Edwards Social Desirability Scale as a Short Form of the MMPI - - - - -	503
<i>Epstein, Seymour.</i> See <i>Lebow, Kenneth E.</i>	
<i>Epstein, Seymour.</i> See <i>Rothstein, Ralph.</i>	
<i>Exner, John E., Jr., McDowell, Eugene, Pabst, Joan, Stackman, William, and Kirk, Lynn.</i> On the Detection of Willful Falsifications in the MMPI - - - - -	91
<i>Eysenck, H. J.</i> Psychoticism or Ten Psychotic Syndromes? - - - - -	179
<i>Fairweather, George W., and Simon, Ralph.</i> A Further Follow-Up Comparison of Psychotherapeutic Programs - - - - -	186
<i>Farber, Norman.</i> See <i>Costa, Louis D.</i>	
<i>Feifel, Herman, and Eells, Janet.</i> Patients and Therapists Assess the Same Psychotherapy	310
<i>Feldstein, Stanley, and Jaffe, Joseph.</i> Language Predictability as a Function of Psychotherapeutic Interaction - - - - -	123
<i>Ferreira, Antonio.</i> See <i>Winter, William D.</i>	
<i>Feshbach, Norma.</i> See <i>Feshbach, Seymour.</i>	
<i>Feshbach, Seymour, Singer, Robert D., and Feshbach, Norma.</i> Effects of Anger Arousal and Similarity upon the Attribution of Hostility to Pictorial Stimuli - - - - -	248
<i>Fishbein, Gerald M.</i> Perceptual Modes and Asthmatic Symptoms: An Application of Witkin's Hypothesis - - - - -	54
<i>Fisher, Jerome.</i> Method for the Study of Staff Interaction through Ratings of Patients -	453
<i>Fisher, Seymour.</i> A Further Appraisal of the Body Boundary Concept - - - - -	62
<i>Fox, Robert, and Lippert, Walter S.</i> Spontaneous GSR and Anxiety Level in Sociopathic Delinquents - - - - -	368
<i>Friedman, Samuel H.</i> See <i>Johannsen, Waller J.</i>	
<i>Gardiner, Sprague H.</i> See <i>Zuckerman, Marvin.</i>	
<i>Garfield, Sol. L.</i> See <i>Helper, Malcolm.</i>	
<i>Ginott, Haim G., and Lebo, Dell.</i> Ecology of Service - - - - -	450
<i>Gofstein, Arnold G.</i> See <i>Dibner, Andrew S.</i>	
<i>Goldstein, Iris Balshan.</i> A Comparison between Taylor's and Freeman's Manifest Anxiety Scales - - - - -	466
<i>Golighly, Carole.</i> See <i>Byrne, Donn.</i>	
<i>Gough, Harrison G., and Domino, George.</i> The D 48 Test as a Measure of General Ability among Grade School Children - - - - -	344
<i>Granick, Samuel.</i> Comparative Analysis of Psychotic Depressives with Matched Normals on Some Untimed Verbal Intelligence Test - - - - -	439
<i>Grebstein, Lawrence C.</i> Relative Accuracy of Actuarial Prediction, Experienced Clinicians, and Graduate Students in a Clinical Judgment Task - - - - -	127



<i>Guyell, Irvin.</i> See <i>Kanfer, Frederick H.</i>	
<i>Gwynn, William.</i> See <i>Slechta, Joan.</i>	
<i>Gynther, Malcolm D.</i> See <i>Brilliant, Patricia J.</i>	
<i>Gynther, Malcolm D.</i> See <i>McDonald, Robert L.</i>	
<i>Haas, Walter A.</i> Investigation of the Stability of Color Preferences - - - - -	537
<i>Hannum, Thomas E.</i> See <i>Miller, William G.</i>	
<i>Hardyck, Curtis, and Petrinovich, Lewis F.</i> The Pattern of Intellectual Functioning in Parkinson Patients - - - - -	548
<i>Hartup, Willard W., Moore, Shirley G., and Sager, Glen.</i> Avoidance of Inappropriate Sex- Typing by Young Children - - - - -	467
<i>Hawkins, L. G.</i> See <i>Caine, T. M.</i>	
<i>Heberlein, Marjorie, and Marcuse, F. I.</i> Personality Variables in the DAP - - - - -	461
<i>Heilbrun, Alfred B., Jr.</i> Evidence Regarding the Equivalence of Ipsative and Normative Personality Scales - - - - -	152
<i>Heilburn, Alfred B., Jr.</i> Revision of the MMPI K Correction Procedure for Improved De- tection of Maladjustment in a Normal College Population - - - - -	161
<i>Heller, Kenneth, Myers, Roger A., and Kline, Linda Vikan.</i> Interviewer Behavior as a Func- tion of Standardized Client Roles - - - - -	117
<i>Helper, Malcolm M., Wilcott, R. C., and Garfield, Sol L.</i> Effects of Chlorpromazine on Learning and Related Processes in Emotionally Disturbed Children - - - - -	1
<i>Hershiser, David.</i> See <i>Pishkin, Vladimir.</i>	
<i>Hesterly, S. O.</i> Deviant Response Patterns as a Function of Chronological Age - - - - -	210
<i>Higashimachi, Wilfred H.</i> The Construct Validity of the Progressive Matrices as a Measure of Superego Strength in Juvenile Delinquents - - - - -	413
<i>Hurst, John G., and Olson, Gordon.</i> Dimensions of Personality Adjustment in Mental Patients - - - - -	465
<i>Hull, S. F.</i> See <i>Crookes, T. G.</i>	
<i>Jaffe, Joseph.</i> See <i>Feldstein, Stanley.</i>	
<i>Johannsen, Walter J., Friedman, Samuel H., Leitschuh, Thomas H., and Ammons, Helen.</i> A Study of Certain Schizophrenic Dimensions and Their Relationship to Double Alter- nation Learning - - - - -	375
<i>Johnson, Gunner.</i> See <i>Smith, Gudmund J. W.</i>	
<i>Jordan, Edward J., Jr.</i> MMPI Profiles of Epileptics: A Further Evaluation - - - - -	267
<i>Kanfer, Frederick H., Bass, Bernard M., and Guyell, Irvin.</i> Dyadic Speech Patterns, Orienta- tion, and Social Reinforcement - - - - -	199
<i>Kaplan, Martin F., and Singer, Erwin.</i> Dogmatism and Sensory Alienation: An Empirical Investigation - - - - -	486
<i>Karp, Stephen A.</i> Field Dependence and Overcoming Embeddedness - - - - -	294
<i>Kennedy, Wallace A., Van De Riet, Vernon, and White, James C., Jr.</i> Use of the Terman- Merrill Abbreviated Scale on the 1960 Stanford-Binet Form L-M on Negro Elementary School Children of the Southeastern United States - - - - -	456
<i>Kirby, Albert C.</i> See <i>Adams, Henry E.</i>	
<i>Kirk, Lynn.</i> See <i>Exner, John E., Jr.</i>	
<i>Klett, C. James.</i> See <i>Lorr, Maurice.</i>	
<i>Kline, Linda Vikan.</i> See <i>Heller, Kenneth.</i>	
<i>Kogan, William.</i> See <i>Boe, Erling E.</i>	
<i>Korman, Maurice, and Blumberg, Stanley.</i> Comparative Efficiency of Some Tests of Cerebral Damage - - - - -	303
<i>Kosowski, Irene.</i> See <i>Walters, Richard H.</i>	
<i>Krieger, Margery H., and Schultz, Charles B.</i> Reliability of Measurement under Various Ex- perimental Instructions - - - - -	459

<i>Kurz, Ronald B.</i> Relationship between Time Imagery and Rorschach Human Movement Responses - - - - -	273
<i>L'Abate, Luciano.</i> See <i>Caputo, Daniel V.</i>	
<i>LaGrone, C. W.</i> Sex and Personality Differences in Relation to Fantasy - - - - -	270
<i>Lane, Ellen A.</i> See <i>Albee, George W.</i>	
<i>Lasky, Julian J.</i> See <i>Lorr, Maurice.</i>	
<i>Lauterbach, Carl G.</i> See <i>Vogel, William.</i>	
<i>Lebo, Dell.</i> See <i>Ginott, Haim G.</i>	
<i>Lebow, Kenneth E., and Epstein, Seymour.</i> Thematic and Cognitive Responses of Good Premorbid Schizophrenics to Cues of Nurturance and Rejection - - - - -	24
<i>Leitschuh, Thomas H.</i> See <i>Johannsen, Walter J.</i>	
<i>Lerner, Barbara.</i> See <i>Cartwright, Rosalind Dymond.</i>	
<i>Levita, Eric.</i> See <i>Costa, Louis D.</i>	
<i>Levy, Bernard I., Lomax, James V., Jr., and Minsky, Raphael.</i> An Underlying Variable in the Clinical Evaluation of Drawings of Human Figures - - - - -	508
<i>Lewinsohn, Peter M.</i> Use of the Shipley-Hartford Conceptual Quotient as a Measure of Intellectual Impairment - - - - -	444
<i>Lippert, Walter S.</i> See <i>Fox, Robert.</i>	
<i>Lomax, James V., Jr.</i> See <i>Levy, Bernard I.</i>	
<i>London, Perry.</i> See <i>Schulman, Robert E.</i>	
<i>Lorr, Maurice, McNair, Douglas M., Klett, C. James, and Lasky, Julian J.</i> Canonical Variates and Second-Order Factors: A Reply - - - - -	180
<i>Lorr, Maurice.</i> See <i>McNair, Douglas M.</i>	
<i>McDonald, Robert L., and Gynther, Malcolm D.</i> MMPI Differences Associated with Sex, Race, and Class in Two Adolescent Samples - - - - -	112
<i>McDowell, Eugene.</i> See <i>Exner, John E., Jr.</i>	
<i>Machir, Daniel F., and Russell, David L.</i> Rater Reliability and Prediction of Diagnosis with the Whittenborn Psychiatric Rating Scales - - - - -	546
<i>McNair, Douglas M., Lorr, Maurice, and Callahan, Daniel M.</i> Patient and Therapist Influences on Quitting Psychotherapy - - - - -	10
<i>McNair, Douglas M.</i> See <i>Lorr, Maurice.</i>	
<i>Marcuse, F. I.</i> See <i>Heberlein, Marjorie.</i>	
<i>Maslow, Harriet I.</i> See <i>Parsons, Oscar A.</i>	
<i>Medinnus, Gene R., and Curtis, Floyd J.</i> The Relation between Maternal Self-Acceptance and Child Acceptance - - - - -	542
<i>Mednick, Martha T.</i> Research Creativity in Psychology Graduate Students - - - - -	265
<i>Meer, Bernard, and Amon, Albert H.</i> Photos Preference Test (PPT) as a Measure of Mental Status for Hospitalized Psychiatric Patients - - - - -	283
<i>Miller, William G., and Hannum, Thomas E.</i> Characteristics of Homosexually Involved Incarcerated Females - - - - -	277
<i>Minsky, Raphael.</i> See <i>Levy, Bernard I.</i>	
<i>Mintz, Elizabeth E.</i> Transference in Co-Therapy Groups - - - - -	34
<i>Moore, Shirley G.</i> See <i>Hartup, Willard W.</i>	
<i>Murstein, Bernard I.</i> The Relationship of Expectancy of Reward to Achievement Performance on an Arithmetic and Thematic Test - - - - -	394
<i>Myers, Albert E.</i> Development of Potentially Therapeutic Interpersonal Relations in Schizophrenics through Competition - - - - -	371
<i>Myers, Roger A.</i> See <i>Heller, Kenneth.</i>	
<i>Neiger, Stephen, and Papastergiou, Christos.</i> The Relationship between Depressive Personality Characteristics and Rorschach Card Preference: A Reply to T. E. Smith - - - - -	463
<i>Neuringer, Charles.</i> Effect of Intellectual Level and Neuropsychiatric Status on the Diversity of Intensity of Semantic Differential Ratings - - - - -	280



<i>Nichols, Robert C., and Schnell, Richard R.</i> Factor Scales for the California Psychological Inventory - - - - -	228
<i>Nickols, John.</i> Rorschach Z Scores on Disturbed Subjects - - - - -	544
<i>Noblin, Charles D.</i> See <i>Timmons, Edwin O.</i>	
<i>Norman, Russell P.</i> Level of Aspiration and Social Desirability in Chronic Schizophrenics	40
<i>Norman, Russell P.</i> Need for Social Approval as Reflected on the TAT - - - - -	464
<i>Norman, Warren T.</i> Relative Importance of Test Item Content - - - - -	166
<i>Nurnberger, John I.</i> See <i>Zuckerman, Marvin.</i>	
<i>Olson, Gordon.</i> See <i>Hurst, John G.</i>	
<i>Pabst, Joan.</i> See <i>Exner, John E., Jr.</i>	
<i>Palmer, Robert D.</i> See <i>Dibner, Andrew S.</i>	
<i>Papastergiou, Christos.</i> See <i>Neiger, Stephen.</i>	
<i>Parsons, Oscar A., Maslow, Harriet I., and Stewart, Kenneth D.</i> Cross-Validation of Stein's Symbol-Gestalt Test for Brain Damage - - - - -	279
<i>Peoples, Crocker.</i> See <i>Slechta, Joan.</i>	
<i>Petrinovich, Lewis F.</i> See <i>Hardyck, Curtis.</i>	
<i>Pierce, Robert C.</i> Note on Testing Conditions - - - - -	536
<i>Pishkin, Vladimir, and Hershisier, David.</i> Respiration and GSR as Functions of White Sound in Schizophrenia - - - - -	330
<i>Pyke, Sandra, and Agnew, Neil McK.</i> Digit Span Performance as a Function of Noxious Stimulation - - - - -	281
<i>Ramer, John.</i> The Rorschach Barrier Score and Social Behavior - - - - -	525
<i>Ransom, Robert.</i> See <i>Winter, William D.</i>	
<i>Raskin, Allen, and Cylde, Dean H.</i> Factors of Psychopathology in the Ward Behavior of Acute Schizophrenics - - - - -	420
<i>Raskin, Allen, and Sullivan, Paul D.</i> Factors Associated with Interrater Discrepancies on a Psychiatric Rating Scale - - - - -	547
<i>Reed, Homer B. C., Jr., and Reitan, R. M.</i> Intelligence Test Performances of Brain Damaged Subjects with Lateralized Motor Deficits - - - - -	102
<i>Reitan, R. M.</i> See <i>Reed, Homer B. C., Jr.</i>	
<i>Rimm, David.</i> Cost Efficiency and Test Prediction - - - - -	89
<i>Rondberg, Samuel R.</i> See <i>Caputo, Daniel V.</i>	
<i>Rosenthal, Irene.</i> Reliability of Retrospective Reports of Adolescence - - - - -	189
<i>Rosenzweig, Saul.</i> Validity of the Rosenzweig Picture-Frustration Study with Felons and Delinquents - - - - -	535
<i>Rothstein, Ralph, and Epstein, Seymour.</i> Unconscious Self-Evaluation as a Function of Availability of Cues - - - - -	480
<i>Russell, David L.</i> See <i>Machir, Daniel F.</i>	
<i>Russell, Peter D., and Snyder, William U.</i> Counselor Anxiety in Relation to Amount of Clinical Experience and Quality of Affect Demonstrated by Clients - - - - -	358
<i>Sager, Glen.</i> See <i>Hartup, Williard W.</i>	
<i>Schaie, K. Warner.</i> Scaling the Scales: Use of Expert Judgment in Improving the Validity of Questionnaire Scales - - - - -	350
<i>Schnell, Richard R.</i> See <i>Nichols, Robert C.</i>	
<i>Schulman, Robert E., and London, Perry.</i> Hypnotic Susceptibility and MMPI Profiles - -	157
<i>Schultz, Charles B.</i> See <i>Krieger, Margery H.</i>	
<i>Scott, Thomas R., Bragg, Robert A., and Smarr, Roy G.</i> Brain Damage Diagnosis with the MMG - - - - -	45
<i>Sechrest, Lee.</i> See <i>Straits, Bruce C.</i>	
<i>Sechrest, Lee.</i> See <i>Wallace, John.</i>	
<i>Shontz, Franklin C.</i> See <i>Tiffany, Donald W.</i>	

<i>Sidle, Allan.</i> Originality in Problem Solving as a Function of Anxiety and Withdrawal in Schizophrenics - - - - -	550
<i>Siller, Jerome, and Chipman, Abram.</i> Response Set Paralysis: Implications for Measurement and Control - - - - -	432
<i>Silver, Albert W.</i> TAT and MMPI Psychopath Deviant Scale Differences between Delinquent and Nondelinquent Adolescents - - - - -	370
<i>Silverman, Jerome.</i> The Validity of the Barron Ego Strength Scale in an Individual Form	532
<i>Simon, Ralph.</i> See <i>Fairweather, George W.</i>	
<i>Singer, Erwin.</i> See <i>Kaplan, Martin F.</i>	
<i>Singer, Robert D.</i> See <i>Feshbach, Seymour.</i>	
<i>Sisson, Boyd D.</i> See <i>Wylie, Ruth C.</i>	
<i>Slechts, Joan, Gwynn, William, and Peoples, Crocker.</i> Verbal Conditioning of Schizophrenics and Normals in a Situation Resembling Psychotherapy - - - - -	223
<i>Smarr, Roy G.</i> See <i>Scott, Thomas R.</i>	
<i>Smith, Gudmund J. W., and Johnson, Gunner.</i> Experimental Prediction of the Long-Term Effect of Psychiatric Therapy - - - - -	373
<i>Snyder, William U.</i> See <i>Russell, Peter D.</i>	
<i>Stackman, William.</i> See <i>Exner, John E., Jr.</i>	
<i>Stewart, Kenneth D.</i> See <i>Parsons, Oscar A.</i>	
<i>Soler, Norton.</i> Client Likability: A Variable in the Study of Psychotherapy - - - - -	175
<i>Straits, Bruce C., and Sechrest, Lee.</i> Further Support of Some Findings about the Characteristics of Smokers and Nonsmokers - - - - -	282
<i>Strickland, Bonnie R., and Crowne, Douglas P.</i> Need for Approval and the Premature Termination of Psychotherapy - - - - -	95
<i>Sullivan, Paul D.</i> See <i>Raskin, Allen.</i>	
<i>Taulbee, Earl.</i> See <i>Wylie, Ruth C.</i>	
<i>Thorne, Gaylord L.</i> Discrimination within the Delinquency Continuum on Gough's Socialization Scale - - - - -	183
<i>Tiffany, Donald W., and Shontz, Franklin C.</i> Fantasized Danger as a Function of Parent-Child Controlling Practices - - - - -	278
<i>Timmons, Edwin O., and Noblin, Charles D.</i> The Differential Performance of Orals and Anals in a Verbal Conditioning Paradigm - - - - -	383
<i>Uyeno, Ensley.</i> Differentiating Psychotics from Organics on the Minnesota Percepto-Diagnostic Test - - - - -	462
<i>Van De Riet, Vernon.</i> See <i>Kennedy, Wallace A.</i>	
<i>Vandiver, James M.</i> See <i>Zuckerman, Marvin.</i>	
<i>Vaughan, Herbert G., Jr.</i> See <i>Costra, Louis D.</i>	
<i>Vestre, Norris D.</i> See <i>Elvekrog, Maurice O.</i>	
<i>Vestre, Norris D.</i> See <i>Wimsatt, William R.</i>	
<i>Vogel, John L.</i> Failure to Validate the Cr and Sm Scales of the MMPI - - - - -	367
<i>Vogel, William, and Lauterbach, Carl G.</i> Sibling Patterns and Social Adjustment among Normal and Psychiatrically Disturbed Soldiers - - - - -	236
<i>Walker, Jerald N.</i> Social Desirability: A Reply to Wiggins - - - - -	458
<i>Wallace, John, and Sechrest, Lee.</i> Frequency Hypothesis and Content Analysis of Projective Techniques - - - - -	387
<i>Wallach, Martin S.</i> Dream Report and Some Psychological Concomitants - - - - -	549
<i>Wallach, Martin S.</i> Judgments of Motivation for Psychotherapy: Some Further Explorations - - - - -	185
<i>Walters, Richard H., and Kosowski, Irene.</i> Symbolic Learning and Reading Retardation - - - - -	75
<i>Waskow, Irene E.</i> Counselor Attitudes and Client Behavior - - - - -	405
<i>Weiss, Janis H.</i> Effect of Professional Training and Amount and Accuracy of Information on Behavioral Prediction - - - - -	257



<i>Werneke, Ann.</i> See <i>Albee, George W.</i>	
<i>White, James C., Jr.</i> See <i>Kennedy, Wallace A.</i>	
<i>Wiggins, Jerry S.</i> Social Desirability under Role Playing Instructions: A Reply to Walker	107
<i>Wilcott, R. C.</i> See <i>Helper, Malcolm M.</i>	
<i>Willingham, Warren W., and Ambler, Rosalie K.</i> The Relation of the Gordon Personal Inventory to Several External Criteria - - - - -	460
<i>Wimsatt, William R., and Vestre, Norris D.</i> Extraexperimental Effects in Verbal Conditioning - - - - -	400
<i>Winter, William D., Ferreira, Antonio J., Ransom, Robert.</i> Two Measures of Anxiety: A Validation - - - - -	520
<i>Wyllie, Ruth C., Sisson, Boyd D., and Taulbee, Earl.</i> Intraindividual Consistency in "Creative" and "Memory" Stories Written for TAT Pictures - - - - -	145
<i>Zuckerman, Marvin, Nurnberger, John I., Gardiner, Sprague H., Vandiver, James M., Barrell, Beatrice H., and den Breeijen, Arie.</i> Psychological Correlates of Somatic Complaints in Pregnancy and Difficulty in Childbirth - - - - -	324





# JOURNAL OF CONSULTING PSYCHOLOGY

February 1963

Vol. 27, No. 1

## Contents

### ARTICLES

- Effects of Chlorpromazine on Learning and Related Processes in Emotionally Disturbed Children: *Malcolm M. Helper, R. C. Wilcott, and Sol L. Garfield* - - - - - 1
- Patient and Therapist Influences on Quitting Psychotherapy: *Douglas M. McNair, Maurice Lorr, and Daniel M. Callahan* - - - - - 10
- A Two-Dimensional Semantic Differential: *Roger B. Allison, Jr.* - - - - - 18
- Thematic and Cognitive Responses of Good Premorbid Schizophrenics to Cues of Nurture and Rejection: *Kenneth E. Lebow and Seymour Epstein* - - - - - 24
- Transference in Co-Therapy Groups: *Elizabeth E. Mintz* - - - - - 34
- Level of Aspiration and Social Desirability in Chronic Schizophrenics: *Russell P. Norman* 40
- Brain Damage Diagnosis with the MMG: *Thomas R. Scott, Robert A. Bragg, and Roy G. Smarr* - - - - - 45
- Perceptual Modes and Asthmatic Symptoms: An Application of Witkin's Hypothesis: *Gerald M. Fishbein* - - - - - 54
- Manifest Anxiety, Social Desirability, or Response Set: *Henry E. Adams and Albert C. Kirby* - - - - - 59
- A Further Appraisal of the Body Boundary Concept: *Seymour Fisher* - - - - - 62
- Symbolic Learning and Reading Retardation: *Richard H. Walters and Irene Kosowski* - 75
- The Use of an Open-Ended Group in the Intake Procedure of a Mental Hygiene Unit: *Andrew S. Dibner, Robert D. Palmer, Benjamin Cohen, and Arnold G. Gofstein* - - 83

### NOTES AND COMMENTS

- Cost Efficiency and Test Prediction: *David Rimm* - - - - - 89
- On the Detection of Willful Falsifications in the MMPI: *John E. Exner, Jr., Eugene McDowell, Joan Pabst, William Stackman, and Lynn Kirk* - - - - - 91

## Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specification given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

### Specifications

**Brief Report.** The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the

author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

<sup>1</sup> An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Document No. — from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$— for microfilm or \$— for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

**Extended report.** Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (a) The extended report should be typed with single spacing for economy in duplication. (b) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (c) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.



## EFFECTS OF CHLORPROMAZINE ON LEARNING AND RELATED PROCESSES IN EMOTIONALLY DISTURBED CHILDREN<sup>1</sup>

MALCOLM M. HELPER, R. C. WILCOTT,<sup>2</sup> AND SOL L. GARFIELD

*Nebraska Psychiatric Institute*

39 children in a psychiatric hospital were tested on 2 types of learning tasks and several performance tasks before, during, and after a 4-week period on chlorpromazine or placebo, the types of medication being assigned randomly in a double-blind design. Medication was controlled according to clinical criteria. Paired-associate learning was less efficient under chlorpromazine, especially on later learning trials and among initially slower learning Ss. Serial learning and tapping rate showed less consistent trends. Porteus Maze Mental Age scores declined under chlorpromazine, while Q scores on the same test tended to improve. No drug effect was noted on remote or immediate memory. The results were tentatively interpreted in terms of impairment of attention span by chlorpromazine.

The purpose of this study was to determine whether the tranquilizer, chlorpromazine, changes the ability of emotionally disturbed children to learn and retain new associations, and if possible, to determine what the basis of any such changes might be. No published research on this specific problem could be found, although the widespread use of the drug with children and the importance of learning during childhood make it a problem of considerable importance.

Previous studies of the effects of chlorpromazine on cognitive processes have used adult subjects almost exclusively, and have produced somewhat inconsistent results. Chlorpromazine has been found to impair retention of verbal associations in adult schizophrenics (Vestre, 1961) and to have no effect on a serial learning task (Whitehead & Thune, 1958). Gilgash (1961) reported improvement in IQ scores in adult catatonics following chlorpromazine, while Gibbs, Wilkens, and Lauterbach (1957) found greater IQ score improvement in adult patients treated with placebo than in those treated with chlorpromazine. Two reports (Porteus, 1957; Porteus & Barclay, 1957) indicated significant declines in Porteus Maze perform-

ance in adult patients treated with chlorpromazine, while a third (Judson & MacCasland, 1960) failed to find a significant effect.

The experimental data on chlorpromazine, while inconclusive, raise the possibility of negative effects on learning and performance in children. However, clinical study of the effects of chlorpromazine in children has led to incidental reports of improved learning of school subjects (Freed & Peifer, 1956). In addition, chlorpromazine has been cited as a useful adjunct in psychotherapy with children, apparently with the belief that the drug facilitates the desired change in behavior (Freed, 1957).

This paper reports the results of a controlled, double-blind study of the effects of chlorpromazine on two learning tasks in children hospitalized for emotional disturbance. Effects of the drug on certain performance tests were also studied in the hope that they would help clarify any learning changes that were found.

### METHOD

#### *Design of the Experiment*

The experiment was divided into three phases: A Predrug phase, an On-drug (chlorpromazine or placebo) phase, and a Postdrug phase. Each child was assessed on each experimental measure at the end of each phase. Four-week intervals separated the successive assessments, the subjects being tested just prior to drug or placebo, again at the end of their fourth week on medication, and finally, at the end of the fourth week following

<sup>1</sup> This investigation was supported by Research Grant No. MY 2318 from the National Institute of Mental Health of the United States Public Health Service.

<sup>2</sup> Now at the Psychology Department, Western Reserve University.

termination of medication. At least a week prior to the first assessment, the children were given a kind of dry run of the assessment procedures to familiarize them with the experimenters and the settings in which the assessments were made.

After being selected for participation in the study (see below), children were assigned randomly to the drug or placebo condition by one of us (RCW) who had no clinical contact with the children and who did not participate in making any of the behavioral measurements. Assignments were made in such a way that the two groups were closely matched for chronological age. Medication was administered and controlled by a psychiatrist who did not otherwise participate in the project or in the administration of the children's activities.<sup>3</sup>

The procedure for administration of the drug was the same as that normally used in clinical practice. Dosage level was adjusted individually for each subject with the initial dosage level being determined by the weight and age of the subject. Dosage level was then increased or decreased as indicated by clinical changes such as drowsiness, decreased motor activity, or other signs of extrapyramidal involvement. The resultant dosage levels ranged from 75 milligrams to 450 milligrams per day. The dosage level was divided into three equal amounts and administered three times a day in liquid form. Placebo dosage level was varied in a similar way. All subjects were examined each week for evidence of drug sensitivity, endocrine disturbance, or jaundice.

### Subjects

All children admitted to the Children's Inpatient Service of the Nebraska Psychiatric Institute between June 1958 and June 1961, were considered for the project. The subjects were excluded if they fell into one of the following categories: (a) Under 6 or over 13 years of age; (b) under 6 years of mental age; (c) too disturbed emotionally to cooperate with testing procedures; (d) judged by the senior psychiatrist to have symptoms which counterindicated the use of chlorpromazine (convulsive disorder or depressive features, for the most part). Five children had to be replaced because of toxic reactions to the drug, and four others had to be replaced because they became physically ill from other causes during the project.

In all, 39 children provided data for the study, 19 drug subjects and 20 placebo subjects. Both groups averaged 115 months in age. The drug group ranged from 86 months to 157 months and contained 15 boys and 4 girls. The placebo group ranged from 81 to 158 months in age and contained 16 boys and 4 girls. The psychiatric diagnoses for the drug group were as follows: Adjustment reaction of childhood, 6; psychoneurotic disorders,

5; schizophrenic reaction, childhood type, 5; personality trait disturbance, 2; chronic brain syndrome with psychosis, 1. Diagnoses for the placebo group were: Adjustment reaction of childhood, 10; psychoneurotic disorders, 7; schizophrenic reaction, childhood type, 2; personality trait disturbance, 1. Although the drug group contains more children diagnosed as psychotic, the two groups initially were not significantly different on any of 18 behavioral scales rated by ward personnel. These scales, which are being reported on elsewhere (Garfield, Helper, Wilcott, & Muffly, 1962), provide a reliable estimate of severity of disturbance. In view of the known unreliability of diagnosis, the two groups can probably be considered to be about equally disturbed.

### Learning Tasks

These tasks had to be selected and designed to secure cooperation from emotionally disturbed children, as well as to yield relevant information. For this reason, new learning tasks were devised which were thought to have more inherent appeal to children than those typically studied in the psychological laboratory. Similarly, more leeway was allowed the experimenter in obtaining and maintaining rapport than would be necessary in dealing with normal children.

*Paired-associate learning.* This task required the subject to associate a picture of a child with a name. The pictures, obtained from magazine illustrations, were reproduced photographically in a standard black and white format and presented on cards with the associated name printed below the picture. The printed name was covered by a folding flap which could be turned down to present the picture and the name together. The names were chosen from a list of popular first names in order to minimize response difficulty. Three eight-item forms of the task were developed, one for each phase of the experiment. During the first phase, pictures of boys and boys' names were used; during the second phase, pictures of girls and girls' names were used; and during the final phase, different pictures of boys and different boys' names were used.

In administering this task, the experimenter instructed the subject to remember the names of the boys (girls). On each presentation, the experimenter both spoke the name aloud and revealed it in printed form after the subject had either made a guess or indicated he did not know. This was done whether or not the subject's response was correct. The pictures were presented in a different prescheduled random order on successive trials. The child was rewarded at the end of each trial with an "M & M" candy for each correct response during that trial. Items were eliminated from the series after two successive correct anticipations. This was done in order to equalize level of learning, since retention was to be tested during the subsequent phase of the experiment. Learning

<sup>3</sup> The authors wish to express their gratitude to Robert B. Muffly, for supervising the administration of the experimental medications.



trials were continued until all items were eliminated or the subject refused further cooperation.

*Serial learning.* Ten 4 × 6 inch cards containing line drawing of familiar objects were presented manually in a constant serial order with instructions to "guess which picture comes next." The correct response was called to the child's attention by the experimenter whenever the child made an error. The series was repeated until the child had reached a criterion of two perfect trials or until the child refused further cooperation. A reward of one small piece of candy (M & M) was given at the end of each trial for each new correct response during that trial. Three different sets of the drawings were used, a different set at each phase of the experiment.

### *Retention Measures*

*Retention of learning tasks.* During the second (On-drug) assessment, the child relearned the series and paired-associate tasks that he had originally learned 4 weeks earlier during the Predrug assessment. This relearning was done prior to the administration of the new forms of the two types of tasks. Similarly, on Postdrug assessment, the child relearned the serial and paired-associate materials that he had originally learned during the On-drug assessment. At this last assessment, the child was not asked to relearn materials originally learned during the Predrug assessment.

*Digit span.* Immediate memory for digit series forward and backward was tested in the usual way, utilizing digit series from the WISC, the WAIS, and the Mental Examiners' Handbook (Wells & Ruesch, 1945) for successive assessments.

### *Performance Measures*

*Tapping rate.* The subjects were asked to tap a stylus as rapidly as possible on a metal plate for three 15-second periods. The taps were counted electrically.

*Dotting O's.* The subjects were asked to place a dot in the center of as many O's on a mimeographed sheet as possible during a 1-minute period. This task requires somewhat finer coordination and discrimination than tapping rate.

*Porteus Maze Test.* Two forms of this test were utilized: The Vineland Revision was administered for the Predrug and Postdrug assessments, while the Extension Series was used for the On-drug assessment.

The first two tasks were included in order to detect drug effects upon such nonlearning factors as cooperation and coordination. The Porteus Maze Test was included because performance on it has been shown to suffer following chlorpromazine in adult schizophrenics (Porteus, 1957; Porteus & Barclay, 1957).

### *Administration of Tasks*

A research assistant administered the learning tasks, the tapping rate test, and the dotting test.

Relearning of old material always preceded learning of new material, occurring a day earlier. Within any one testing session, learning or relearning of the serial task preceded the corresponding operation on the paired-associate task. Two research assistants were employed over the course of the 3-year period of data collection, each testing equal numbers of drug and control subjects. The Porteus Maze Test and Digit Span were administered by the first author in the presence of the third author, who was involved in making ratings for other aspects of the study.

## RESULTS

### *General Form of Analysis*

Changes between the successive assessments were tabulated separately for the two groups for each measure. Effects of chlorpromazine administration were then studied by comparing the Predrug to On-drug changes in the experimental group with those in the control group. Since there was wide variation among the subjects in initial level on most of the variables, and since subsequent changes appeared to be correlated with initial level, change scores for each variable were entered into an analysis of variance in which the subjects were classified according to initial level for that variable, as well as according to kind of medication. For any given variable, some subjects were lost in the post hoc matching of subgroups on initial level. For this reason, analyses were carried out on each variable ignoring the initial level classification and including all subjects providing relevant data. Only one result changed in significance level under this more inclusive form of analysis (see results for Porteus Qualitative score).

Changes in the two groups following the termination of medication were also compared, following the pattern described above. Predrug values were again used to make the initial level classifications. Three subjects in the control group were discharged prior to the third assessment, reducing the number available for comparisons of Postdrug changes.

### *Learning and Retention Measures*

*Paired-associate learning.* Only the first eight trials of each paired-associate task were included in these analyses, since the number of cooperative subjects dropped after this point in both the Predrug and On-drug

TABLE 1  
VARIANCE ANALYSES OF CHANGES IN NUMBER CORRECT ON PAIRED-ASSOCIATE LEARNING

Source	Predrug to On-drug change			On-drug to Postdrug change		
	<i>df</i>	<i>MS</i>	<i>F</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Drug (D)	1	62.68	4.48*	1	30.99	
Level (L)	1	.56		1	33.01	
D × L	1	25.84		1	8.98	
Between subjects error	32	14.00		28	13.14	
Trial block (T)	3	1.21		3	3.13	
T × D	3	15.63	6.25**	3	8.60	2.86*
T × L	3	1.00		3	3.76	
T × D × L	3	16.69	6.68**	3	1.80	
Within subjects error	96	2.50		84	3.00	
Total	143			127		

\*  $p < .05$ .

\*\*  $p < .01$ .

assessments. On the Predrug assessment, 20 of the 39 subjects had reached the criterion as of the eighth trial, and the average number of correct responses on the eighth trial was 6.7, out of a possible 8. These data indicate that a substantial amount of learning had occurred by the eighth trial.

The eight trials were divided into four blocks of two trials each. This was done so that drug effects upon the course of learning could be studied. Changes between the Predrug and On-drug tasks were assessed by subtracting the number correct on each Predrug block from that in the corresponding On-drug block. On-drug to Postdrug changes were tabulated analogously, after subtracting the former from the latter.

In the analyses of variance performed on these two sets of change scores, the following effects were studied: Drug (chlorpromazine or placebo); initial learning level (the subjects were divided at the grand median on total number of errors on the Predrug task); and trials (four two-trial blocks). These analyses are summarized in Table 1. Thirty-six subjects, divided into four subgroups of 9 each, were utilized in the analysis of Predrug to On-drug changes. Only 28 subjects could be used in the analysis of On-drug to Postdrug changes. Three control subjects were discharged from the hospital before the final assessment, and only 14 of the remaining 16 controls could be matched with experimental subjects on Predrug learning scores.

It can be seen that treatment (Drug), the Drug by Trial interaction, and the triple interaction all represent significant effects in the Predrug to On-drug analysis. The changes in number correct on the four trial blocks are plotted separately for the four subgroups in Figure 1. These changes seem to permit a fairly simple interpretation: Chlorpromazine results in a decrement in correct responses, especially on later learning trials, and most especially in subjects who were slow learners prior to its administration. The analogous analysis of Postdrug change scores, shown in the right half of Table 1, resulted in a significant Drug by Trial interaction. Examination of the relevant mean Postdrug changes, plotted in Figure 2, points to an interpretation which partially parallels that above—upon termination of treatment, the chlorpromazine group showed a large recovery in performance on the later learning trials, especially as compared with the control group. Taken together, the analyses point to a decline under chlorpromazine of performance on later trials in the paired-associate learning task, a loss from which significant recovery was made within 4 weeks after termination of the drug. In addition, there is some suggestion that slower learning subjects are affected more by chlorpromazine.

*Serial learning.* On the Predrug assessment, 15 of the 39 subjects had reached the criterion by the eighth trial of this task, and



the mean number of correct responses on this trial was 7.9 out of a possible 9. These data indicate that the serial task was roughly comparable in difficulty level to the paired-associate task. Changes in number correct on the serial tasks were also tabulated for four two-trial blocks. Three-way analyses of variance were performed on the Predrug to On-drug changes and on the On-drug to Postdrug changes, just as with the paired-associate task. Initial level classifications were made on the basis of total number correct on the first eight trials on the Predrug serial task. No significant effects involving chlorpromazine emerged from the Predrug to On-drug analysis. However, two interactions involving type of medication were significant in the analysis of Postdrug changes. Because these rather complex changes upon removal of medication would seem to have little meaning in the absence of any significant changes with initial drug administration, these interactions will not be reported in detail.

*Retention of learned materials.* One form of each learning task was originally learned prior to medication and relearned 4 weeks later under drug or placebo. The second form of each task was learned originally under drug or placebo and relearned 4 weeks later after termination of medication. Two retention scores were derived and were used with each form of each type of task: (a) A *recall* score consisting simply of the number of correct responses on the first relearning trial; and (b) a *percentage savings* score, computed by subtracting the number of errors on relearning from the number of errors on original learning and dividing the remainder by the latter term. Separate analyses of variance were performed on the two retention scores for each form of each task, subcategorizing subjects on the basis of the Predrug learning of the task involved. No significant effect on either retention score for either type of task was associated with either the introduction or removal of chlorpromazine.

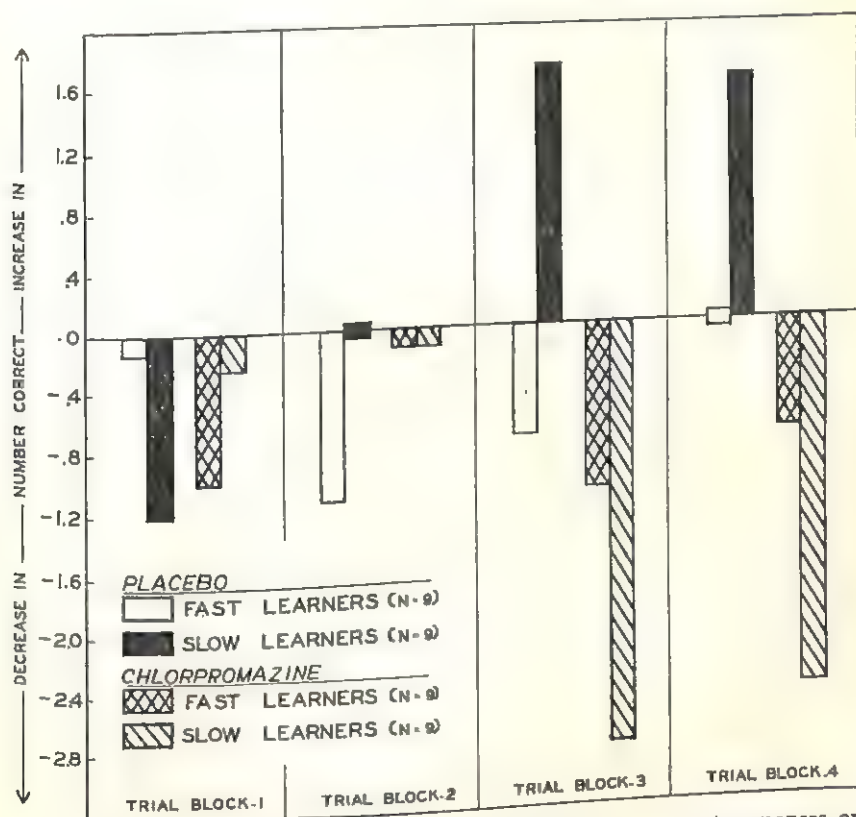


FIG. 1. Predrug to On-drug changes in mean number of correct responses on four two-trial blocks of paired-associate learning. (Changes are shown separately for initially fast and slow learning subjects within each treatment group.)

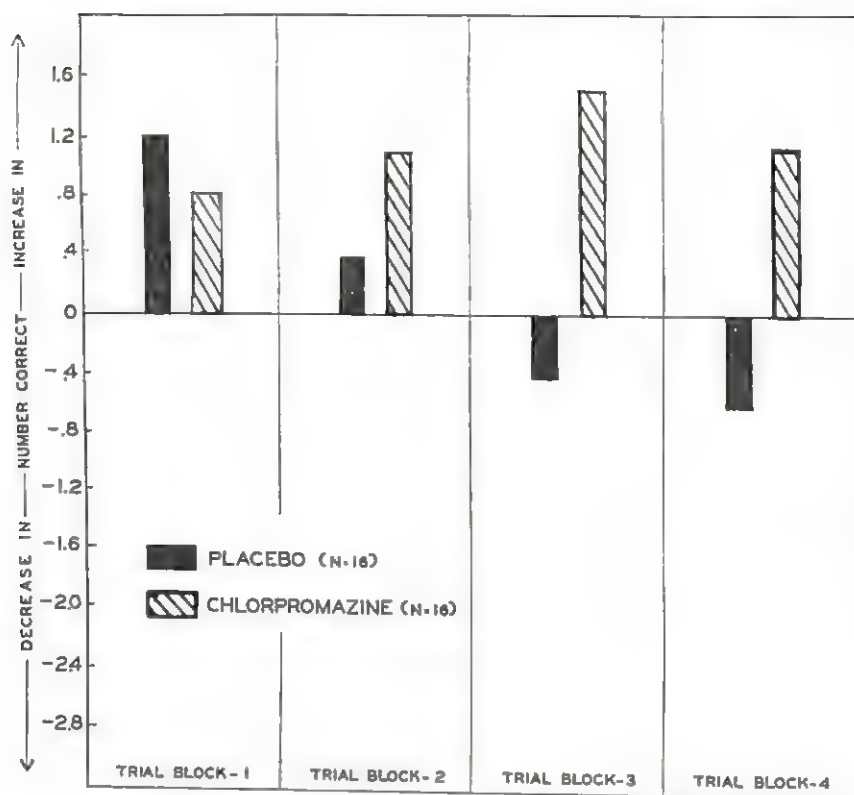


FIG. 2. On-drug to Postdrug changes in mean number of correct responses on four two-trial blocks of paired-associate learning.

*Digit Span.* No significant difference between experimental and control groups was found on Digit Span, nor was any tendency toward such a difference discernible.

#### Performance Measures

*Tapping rate.* Each set of change scores for (Predrug to On-drug and On-drug to Postdrug) was entered into a three-way analysis of variance, utilizing Drug, Trials, and Initial Level as main classifications. Initial Level was determined from the total tapping score on the Predrug assessment. Thirty-six subjects could be included in the first analysis, but only 28 in the second, because of the Postdrug loss of subjects.

No significant effects involving medication appeared in the analysis of Predrug to On-drug changes. The one result to reach statistical significance in the analysis of Postdrug changes, an interaction between Trial and Drug, would seem to be of little theoretical or practical import in the absence of any

measurable effect of the initial chlorpromazine treatment.

*Dotting O's.* The other performance task, Dotting O's, revealed no significant effects when analyzed in the same fashion as above.

*Porteus Maze mental age.* Predrug to On-drug changes were entered into an analysis of variance, as were the On-drug to Postdrug changes. In both analyses, the Predrug scores were used to make the Initial Level classifications. Four equal groups, one for each combination of Level and Drug, were used in each analysis. A total of 36 subjects was used in the analysis of Predrug to On-drug changes, while 32 subjects were available for the analysis of Postdrug changes. The results of these analyses, presented in Table 2, show a significant drug effect on the Predrug to On-drug changes (the significant Levels effects are of little interest for the present study, since they do not involve chlorpromazine in any way).

Examination of the mean Predrug to On-drug changes in the two treatment groups



TABLE 2  
VARIANCE ANALYSES OF CHANGES IN PORTEUS MAZE MENTAL AGE

Source	Predrug to On-drug changes			On-drug to Postdrug changes		
	df	MS	F	df	MS	F
Drug (D)	1	10.56	4.84*	1	.07	
Level (L)	1	16.67	7.65**	1	25.38	5.52*
D × L	1	1.56		1	2.26	
Between subjects within groups	32	2.18		28	4.59	
Total	35			31		

\*  $p < .05$ .

\*\*  $p < .01$ .

showed that the experimental group as a whole declined an average of 1.2 years in mental age score, while the control group lost only 0.1 year, a net difference of 1.1. The nonsignificant difference between the groups in posttreatment changes reflects the failure of the chlorpromazine subjects to recover in mental age score as sharply as or as consistently as they lost it with drug administration. However, the chlorpromazine group did recover enough after the termination of the drug so that its net change from its Predrug level was not significantly different from the corresponding change in the control group. (These latter data are not shown in tabular form.)

*Porteus Maze Q score.* This score is derived by counting and weighting various kinds of errors (e.g., crossing lines, cutting corners) that do not enter directly into the mental age scores. It is considered by Porteus to reflect the amount of impulsiveness and carelessness exhibited by the subject. In this study, the total Q score was divided by the number of mazes actually administered to the subject, so that the values analyzed reflected errors per maze attempted. This adjustment was made because the subjects varied widely in ability level and, therefore, in the number of mazes which they attempted before reaching the criterion for cessation of testing. While very little change in the Q score occurred in the control group, it declined following the introduction of chlorpromazine in the experimental group, increasing again after the termination of the drug. The decrease was not quite significant

by comparison with changes in the control group ( $p = .07$ ), but the latter change was significantly different ( $p = .05$ ) from that observed in the placebo group. These results were obtained when categorization of subjects by initial Q score was dropped so that all subjects could be included. It should be noted that a reduction in this error score signifies an improvement in qualitative aspects of maze performance.

#### DISCUSSION

The effects attributable to the administration of chlorpromazine in the present study were selective and specific rather than general and pervasive. Only two of the measures studied, the paired-associate learning task, and the Porteus Maze Mental Age score, changed more with the administration of chlorpromazine than with placebo. Selectivity in drug effects was noted within the paired-associate task itself, in that the later trials and slower learning subjects were most affected by chlorpromazine. The serial learning task, digit span, tapping rate, dotting rate, and two measures of retention of the serial and paired-associate material showed no effects related to the administration of chlorpromazine. Significant interactions involving chlorpromazine did occur for serial learning and tapping rate following the removal of medication. However, these rather complex Postdrug interactions alone, without corroborative effects with initial drug administration, were deemed too unreliable to be useful in generalizing about the influence of chlorpromazine.

Although few in number, the significant effects that were observed with the introduction of chlorpromazine were consistent in direction. On both the paired-associate measure and the Porteus Mental Age score, the subjects who received the active drug declined in performance relative to those who received placebo. Consistency of response was further noted in the paired-associate task in that the decrement associated with the administration of chlorpromazine was reversed following removal of medication.

Since previous experimental work on the cognitive effects of chlorpromazine has apparently not used emotionally disturbed children as subjects, the present results will have to be compared with the most nearly similar studies with adults. The findings for the paired-associate task in the present study are in only partial agreement with those obtained from adult schizophrenics by Vestre (1961) in a study of the acquisition of non-preferred word associations. He found a trend toward less efficient learning of the task as a whole in the chlorpromazine subjects, but this was not statistically significant. He did not examine the later trials separately. Also in contrast to the present study, he found a significant decrement in retention for those materials learned under chlorpromazine. The two studies thus seem to agree in showing negative effects of chlorpromazine, but to differ in the exact nature of these effects. The nonsignificant results for the serial task in the present study agree with those of Whitehead and Thune (1958). Using adult psychotic subjects, these investigators found no difference between chlorpromazine and placebo groups in learning or retention of an eight-item verbal series.

The relative decline of about 1 year in Porteus Maze Mental Age following chlorpromazine in the present study was somewhat less than the average deficit of about 2 years reported by Porteus (1957) and Porteus and Barclay (1957), but does indicate negative effects for chlorpromazine on this test in children as well as in adults. The present study also revealed a trend toward improved qualitative performance (*Q* score) on this test under chlorpromazine. This variable has apparently not been studied be-

fore in experiments on chlorpromazine, though it could have valuable implications for the nature of the deficit produced by chlorpromazine on the Porteus Mental Age score.

The nonsignificant results for digit span, tapping rate, and dotting rate agree with nonsignificant findings on similar variables for adults (Levita, 1961; Shatin, Rockmore, & Funk, 1956).

To sum up, the present results from emotionally disturbed children would seem to correspond in broad outline with those obtained on similar cognitive tasks with adults. In both cases the measurable effects of chlorpromazine are relatively few, but rather consistently negative in direction.

What do these results indicate about the mode of action of chlorpromazine? Two negative conclusions would seem to follow: The data do not support the notion that chlorpromazine will typically improve learning or other cognitive performances in disturbed children, as might be expected if the drug only reduced anxiety or distractibility. At the same time, the specificity of the negative effects observed makes it unlikely that the drug impairs a factor basic to all learning tasks, such as the formation or maintenance of associational bonds. In fact, the concentration of negative drug effects within the latter trial blocks on the paired-associate task suggests that the factor affected by chlorpromazine can vary in importance within the course of a single learning session, as well as from one learning task to the next.

Do the data contain any more positive suggestions as to the nature of the factor impaired by chlorpromazine? In attempting to narrow the range of speculation, we can ask whether the Porteus Mental Age score, which also suffers under chlorpromazine, has anything in common with the paired-associate task, especially with its later trials. The Predrug Porteus Mental Age score was found to correlate significantly ( $r = .44$ ,  $N = 39$ ,  $p = .05$ ) with the number correct on the third trial block of the Predrug paired-associate task. Interestingly enough, the corresponding score for the serial task, which did not show impairment with the adminis-



tration of chlorpromazine, did not correlate significantly with Porteus Mental Age ( $r = .25$ ,  $N = 39$ ). The Porteus  $Q$  score, which seemed to improve under chlorpromazine, has no Predrug correlation in excess of .11 with any of the three variables mentioned above. These data would suggest, though not unequivocally, that chlorpromazine impairs some factor which is shared by the Porteus Mental Age score and the later paired-associate trials, but which is not heavily represented in the corresponding serial learning trials or in the Porteus  $Q$  score. One possibility is that chlorpromazine impairs the maintenance of active attention to novel and significant details. The paired-associate task, with its continued reshuffling and elimination of items, and the Porteus Maze, with its series of increasingly complicated designs, would seem to share some requirement for such sustained alertness. This interpretation is made more attractive by the work of Mirsky and Rosvold (1960), which related lapses of attention on the Continuous Performance Test following chlorpromazine to depression of functioning in the midbrain reticular formation. It is recognized, of course, that other factors than attention may be affected by chlorpromazine and may be more obvious in experiments with other tasks and other subjects.

## REFERENCES

- FREED, H. The tranquilizing drugs and the school child. *Amer. Pract.*, 1957, 8, 377-380.
- FREED, H., & PEIFER, C. Some considerations on the use of chlorpromazine in a child psychiatry clinic. *J. clin. exp. Psychopath.*, 1956, 17, 165-169.
- GARFIELD, S. L., HELPER, M. M., WILCOTT, R. C., MUFFLY, R. B. Effects of chlorpromazine on behavior in emotionally disturbed children. *J. nerv. ment. Dis.*, 1962, 135, 147-154.
- GIBBS, J., WILKENS, B., & LAUTERBACH, C. A controlled clinical psychiatric study of chlorpromazine. *J. clin. exp. Psychopath.*, 1957, 18, 269-283.
- GILGASH, C. A. Thorazine therapy with catatonic schizophrenics in relation to Wechsler verbal and performance subtest comparison. *J. clin. Psychol.*, 1961, 17, 95.
- JUDSON, A. J., & MACCASLAND, B. W. The effects of chlorpromazine on psychological test scores. *J. consult. Psychol.*, 1960, 24, 192.
- LEVITA, E. Effects of chlorpromazine and promazine on perseveration. *J. gen. Psychol.*, 1961, 65, 181-187.
- MIRSKY, A. F., & ROSVOLD, H. E. The use of psychoactive drugs as a neuropsychological tool in studies of attention in man. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 375-392.
- PORTEUS, S. D. Maze test reaction after chlorpromazine. *J. consult. Psychol.*, 1957, 21, 15-21.
- PORTEUS, S. D., & BARCLAY, J. E. A further note on chlorpromazine: Maze reactions. *J. consult. Psychol.*, 1957, 21, 297-299.
- SHATIN, L., ROCKMORE, L., & FUNK, I. C. Response of psychiatric patients to massive doses of thorazine: II. Psychological test performance and comparative drug evaluation. *Psychiat. Quart.*, 1956, 30, 402-416.
- VESTRE, N. D. Effects of thorazine on learning and retention in schizophrenic patients. *J. abnorm. soc. Psychol.*, 1961, 63, 432-435.
- WELLS, F. L., & RUESCH, J. *Mental examiners' handbook*. New York: Psychological Corporation, 1945.
- WHITEHEAD, W. A., & THUNE, L. E. The effects of chlorpromazine on learning in chronic psychotics. *J. consult. Psychol.*, 1958, 22, 379-383.

(Received May 11, 1962)



## PATIENT AND THERAPIST INFLUENCES ON QUITTING PSYCHOTHERAPY

DOUGLAS M. McNAIR, MAURICE LORR, AND DANIEL M. CALLAHAN

*Veterans Benefits Office, Washington, D. C.*

Groups of 106 early therapy Terminators and 176 Remainers differed significantly on 10 patient characteristics. Best predictors of duration were a Terminator-Remainer test battery plus rated motivation for psychotherapy. Therapists with marked interest in the patients' problems held significantly higher proportions of both predicted Terminators and Remainers in treatment. 7 other therapist characteristics did not substantially influence termination rates. Hiler's contention that predicted Terminators and Remainers respond differently to the same therapist attributes was not supported, but it was found that some therapists react differently to the 2 types of patients. The results indicate that Terminators and Remainers form distinguishable out-patient populations. The former reject psychotherapy, perhaps because they lack the behavioral repertoire required for participation.

Many studies in the past decade have aimed at predicting premature termination of psychotherapy. Fulkerson and Barry (1961, pp. 194-197) have presented a review and bibliography of the principal studies of the problem. Lorr, Katz, and Rubinstein (1958) reported the development and cross-validation of a test battery (TR battery) to predict early termination of psychotherapy. Results of a double cross-validation provided substantial evidence that the test battery identified a high proportion of patients who terminated psychotherapy after a few interviews without their therapists' advice or consent. The TR battery predicted termination with significantly greater accuracy than base-rate predictions and, contrary to the statement in Fulkerson and Barry's review, with significantly higher accuracy than predictions by an interviewer. The characteristics assessed by the TR battery indicated,

The Terminator is either not anxious or does not admit to being anxious and self-dissatisfied. He is likely to have a history of antisocial acts, he admits to being undependable and impulsive, and may be authoritarian or rigid in his social attitudes (Lorr et al., 1958, p. 326).

One aim of the present study was to attempt to increase the predictive accuracy of the TR battery and to obtain greater knowledge of Terminator response patterns by including additional patient measures in a study of premature termination of treatment. A corollary

aim was to perform a second cross-validation of the TR battery.

Studies of termination have emphasized patient variables related to dropping out of therapy. Few studies systematically assessed the influence of nonpatient variables on premature termination of therapy, although some investigation suggested the unexplained criterion variance might be due to variables such as therapist behavior and the setting in which treatment takes place. Lorr et al. (1958) reported that therapist sex, profession, and experience did not relate significantly to termination. They also noted that setting factors such as scheduled treatment frequency, length of interview, and type of therapy (group versus individual) did not correlate significantly with the Terminator-Remainer criterion. Recently, however, Hiler (1958) presented evidence that the type of patient terminating related to characteristics of the therapist. Briefly, he reported that therapist sex, warmth, and psychotherapeutic competence were associated with whether therapists retained predicted Quitters or predicted Stayers in treatment. ("Quitter" and "Terminator," "Stayer" and "Remainer" are used synonymously.) He typed his patients as Stayers or Quitters on the basis of total Rorschach responses—an index with some cross-validated evidence. It should be noted that Hiler reported an interaction effect between therapist characteristic and patient type on outcome—not an

association between therapist characteristic and actual termination rate. Thus his findings do not necessarily disagree with Lorr's.

Hiler's (1958) study provided the interesting lead that therapist-patient compatibility influenced premature termination, and his results suggested that the interaction between the therapist and patient "type" might be crucial. His findings, however, were based on small samples from a single clinic and there was no cross-validation of the interaction. Since no data were presented giving the correlations between sex, warmth, and competence, their relative independence could not be judged. Some of Hiler's one-tailed test findings (e.g., competence and patient type) would not be significant if two-tailed tests had been used. One-tailed tests seem questionable in such an unexplored area. Finally, certain relevant data were not presented by Hiler. He reported an interaction of therapist warmth and patient type among Remainers; no comparable test for Terminators was mentioned. He reported the interaction of therapist competence and patient type among his Terminators but cited no comparable test for Remainers. While he reported data relevant only to stated hypotheses, it is unclear why comparable (or converse) hypotheses were not tested for both Terminators and Remainers. Thus a third aim of the current report was to investigate the effect of the therapist on termination and to attempt confirmation of Hiler's interactions between therapist characteristics and patient type.

Terminators were operationally defined as patients who refused or stopped therapy in less than 16 weeks without the advice or consent of their therapists. The study hypotheses, based on surveys of previous findings in the area, were:

1. The TR battery predicts better than chance and better than the base rate those patients who will terminate therapy prematurely.
2. Terminators have more poorly developed verbal behavior than Remainers.
3. Terminators have less formal education—the crucial component of most social class indices (Freeman, 1961)—than Remainers.
4. Terminators admit less anxious and "neurotic" behaviors than Remainers.

5. Compared with Remainers, Terminators have a history of more impulsive and anti-social behavior, more difficulties with the law, and lack close interpersonal relationships.

6. Terminators are less interested than Remainers in participating in verbal psychotherapies.

7. Compared with Remainers, Terminators admit less dissatisfaction with their own characteristics and behavior.

8. Terminators more frequently than Remainers endorse rigid, authoritarian, irrational social attitudes, and opinions.

9. When patients are classified in advance as probable Terminators or Remainers on the basis of a predictively valid test, interaction effects on outcome will appear between patient type and characteristics of the therapist such as sex and competence.

## METHOD

The sample consisted of 282 psychotherapy outpatients in seven Veterans Administration Mental Hygiene Clinics.<sup>1</sup> All patients satisfied these criteria for inclusion in the study: males under age 51; without evidence of CNS damage; less than 12 hours of therapy during the preceding 90 days; acceptable to the clinic for "intensive" psychotherapy—defined as weekly interviews lasting about 1 hour, aimed at changing personal adjustment patterns, and excluding supportive or maintenance therapy. Clinics added research procedures to their customary prepsychotherapy processing routines. The assessment included a research test battery, evaluation by a social worker, and appraisal by the therapist. When a patient terminated prior to the 16-week evaluation date, a data sheet was completed giving the date and reason for termination. There were 176 Remainers and 106 Terminators. Patients who terminated for reality reasons—e.g., moving, requiring hospitalization—were excluded from analysis. Within 6 weeks or less, 74 of the Terminators quit therapy—confirming the observation of Lorr et al. (1958) that most premature terminations occur during the first few weeks of treatment.

Only 2 of the 7 clinics involved were among the 13 in the Lorr et al. (1958) study. All clinics in

<sup>1</sup>We are grateful to the following psychologists and to other staff members at their stations for participating in the study: R. W. Boyd (Boston), L. Hemmendinger (Bridgeport), P. W. Morse (Hartford), L. Rutledge (Denver), J. E. Tucker (Albany), H. H. Weiss (Chicago), and B. B. Yormak (Buffalo).

The data were collected in collaboration with a larger Veterans Administration Cooperative Project (Lorr, McNair, Michaux, & Raskin, 1962).



both studies are located in large cities widely distributed geographically and all clinics had 10-40 staff members. Most therapists in both studies had Freudian or Rogerian orientations. The present data were collected throughout 1957 and 1958; the Lorr et al. data in 1955-56. World War II veterans comprised more than 90% of both samples. The remainder were Korean veterans. A separate study (McNair, Callahan, & Lorr, 1962) of the effect of the therapist on response to psychotherapy involved 40 of the current Remainders.

The Terminator-Remainder criterion in the present study differed from those of Lorr et al. (1958) and Hiler (1958). Lorr's Terminators and Remainders were defined, respectively, as patients remaining in therapy 6 weeks or less and patients remaining 6 months or longer. Hiler's Terminators had 5 or fewer interviews; his Remainders had 20 or more interviews. Both studies excluded the middle group from analysis. Inclusion of the middle group in the current project probably made for a more severe test of the hypotheses, especially Hypothesis 1. Obvious Terminators were probably excluded from the present sample since the criteria for inclusion were more rigorous than the usual clinic criteria for acceptance into psychotherapy. The 16-week termination rate (38%) is much lower than the usual 6-week base rate (about 50%) in Veterans Administration clinics.

### Predictor Variables

Patient predictor variables are briefly described below. The arabic number in parentheses indicates the hypothesis tested by each predictor(s). In addition to those below, 10 other patient measures, which proved unrelated to the Terminator-Remainder criterion, were included as tests of Hypotheses 2-8.

*Patients predictors.* (1) TR-A, TR-BD, and TR-F are, respectively, six-, eight-, and seven-item subtests which constitute the TR Battery. They are taken from the longer Manifest Anxiety (MA) scale, Behavior Disturbance (BD) scale, and F Scale described below; (2) a 30-item multiple-choice vocabulary test adapted from the Cooperative English Test, Form Z, Educational Testing Service; (3) a nine-point scale of formal education completed; (4) a slightly modified form of Taylor's (1953) MA scale and a Sociability scale abbreviated from a scale in the Guilford-Zimmerman (1949) Temperament Survey; (5) BD: 50 questionnaire items adapted from an inventory by Applezweig, Dibner, and Osbourne (1958) to measure antisocial behavior and psychopathic tendencies; (6) Motivation for Treatment: rating by the therapist on a four-point graphic scale of the patient's interest in obtaining psychotherapy; (7) Ideal-Actual Self-Discrepancy: a discrepancy score between ratings of the actual self and the ideal self on 21 five-point scales designed to measure self-satisfaction; (8) F Scale: 25 items from the Adorno, Frenkel-Brunswick, Levinson, and Sanford (1950) measure of authoritarianism.

TABLE 1

DIFFERENCES BETWEEN REMAINERS AND TERMINATORS ON 10 CHARACTERISTICS

Variable	M		t
	Re- main- ers	Ter- mi- nators	
TR-BD	5.9	5.3	3.10**
TR-A	3.3	2.8	2.97**
TR-F	4.0	3.1	3.84**
MA scale	29.6	27.1	2.39*
F Scale	24.5	23.1	3.26**
Sociability	7.3	9.0	3.70**
Ideal-Actual Self	34.6	30.9	2.61**
Education*	5.3	4.8	3.31**
Vocabulary	22.4	20.2	3.44**
Motivation	2.7	2.3	4.93**

\* Score of 5 on education indicates high school graduation.

\*\*  $p < .05$ .

\*\*  $p < .01$ .

*Therapist predictors.* Eight therapist predictors were included to test Hypothesis 9. These were: sex; profession; years of experience as a therapist; amount of personal psychotherapy; competence; liking for the patient; interest in the type of problem presented by the patient; and therapist A or B "type." Competence ratings were obtained by having three clinical psychologists with 5 or more years experience audit tape recordings of psychotherapy sessions conducted by the therapists (interrater agreement was .68<sup>2</sup>). Liking and Interest were rated on graphic scales by therapists at the conclusion of the initial therapy hour. Therapists were typed as A's or B's on the basis of their scores on 23 items from the Strong Vocational Interest Blank. Whitehorn and Betz (1960) found these items discriminated between groups of therapists who obtained different success rates (A's were more successful) with hospitalized schizophrenics. McNair et al. (1962) found that A and B therapists also obtained significantly different results with outpatients (but B's were more successful).

## RESULTS

### Cross-Validation of the TR Battery

Ten patient predictors significantly differentiated Terminators and Remainders (Table 1). All three tests in the TR battery significantly discriminated the two groups of patients. As in previous studies, the TR battery subtests discriminated more clearly than their parent tests. The point biserial correlations of TR-A and TR-F with the criterion were slightly higher than for the MA scale and the F scale, while Terminators and Re-

<sup>2</sup> Interclass correlation coefficient.



TABLE 2  
OUTCOME OF TREATMENT FOR PATIENTS PRE-  
DICTED TO REMAIN AND TERMINATE

Configural prediction	Outcome		Total
	Re- main	Termi- nate	
Remain	119	38	157
Terminate	57	68	125
Total	176	106	282

Note.— $\chi^2 = 27.05$ ;  $df = 1$ ,  $p < .001$ .

mainers did not differ significantly on the total BD inventory. The product-moment correlation between TR-A and MA scale for the entire sample was .83. The correlation is somewhat inflated, as the TR-A items are in the MA scale. Even so, the correlation is as high as the MA scale reliability and indicates TR-A can be considered an abbreviated MA scale.

The configural prediction approach suggested by Lorr et al. (1958) was applied to the pattern of scores on the TR battery. Using cutting scores from the original study, patients' scores on each subtest were classified as either Terminator or Remainer scores. The eight resulting patterns of scores on the three subtests in the battery were then classified as Terminator or Remainer predictor patterns. Table 2 presents the outcome of treatment for patients with the two classes of patterns. Chi square for the accuracy of prediction was highly significant. Overall accuracy of prediction was 66% with 64% of the actual Terminators and 68% of Remainers identified.

Lorr's original beta weights<sup>a</sup> were used to find the multiple correlation ( $R$ ) of the TR battery with the criterion for the current sample. The  $R$  obtained was .44 ( $p < .001$ ), a value that compares favorably with  $R$ s of .67 in Lorr's original sample and .39 in his cross-validation sample. The difference between  $R = .44$  and the beta of .36 between the configural scores and the criterion was tested (Osburn & Lubin, 1957), and  $R$  is significantly larger. The cutting score based on the multiple regression equation yielding maxi-

mum accuracy identified 72% of the Remainers, 64% of the Terminators, with 69% overall accuracy. Compared with the configural approach, use of the multiple regression equation increased accuracy of identifying Remainers with no loss of accuracy among the Terminators.

Use of the configural scores increased accuracy of predicting Remainers 6% above the sample base rate, while  $R$  increased accuracy 10%. Only the latter represented a significant increase ( $\chi^2 = 4.10$ ,  $p < .05$ ). Judgment of the actual utility of the TR battery should consider the facts that these patients had to meet rather rigorous criteria for inclusion, that the criterion groups were defined differently, and that the analysis was not confined to extreme groups. It should also be noted that the TR battery increased accuracy of identifying Terminators by almost the same percentage as Lorr et al. (1958) found. Here the increase was from 38% to 64%; their sample included 55% Terminators, and 80% were accurately identified.

#### *Additional Patient Predictors*

The findings in Table 1 support the hypotheses that Remainers have a history of less impulsive and less antisocial behavior, admit more anxious behavior, are more critical of themselves, and are less likely to endorse rigid, irrational beliefs. They are also more retiring in interpersonal relationships, better educated, have better vocabularies, and therapists consider them more highly motivated for psychotherapy.

Do any combinations of the measures of these characteristics offer potentialities for increasing the accuracy of predicting termination of therapy? The Lubin-Summerfield (1951) square root method was applied to the 10 significant variables in Table 1. The procedure selects the minimum number of variables which yields the highest possible  $R$  with the criterion. Variables are added in order of decreasing proportion of criterion variance accounted for. Five variables were added to the minimum set in the following order: Motivation, TR-BD, TR-A, TR-F, and Sociability. The multiple  $R$  for the five variables with the criterion was .48. The five variables accounted for little more criterion

<sup>a</sup> Beta weights and cutting scores are available upon request.

variance than the TR battery alone, using the original beta weights.

### *Therapist and Duration*

Patients were divided into two groups or types—predicted Quitters and Stayers—on the basis of the duration predicted by applying the multiple regression equation to their TR battery scores. Effects of therapist variables were tested by determining if actual duration for the two groups related to the therapist variable operating. Table 3 presents the proportion and number of predicted Quitters and Stayers who remained in treatment with therapists who differed on eight characteristics. The *N*s in Table 3 are smaller than in the preceding analyses because Terminators who kept no therapy appointments were omitted; *N*s in Table 3 are variable, depending on availability of data.

Statistical analyses of the Table 3 data were similar for all variables. An overall  $\chi^2$  test for the four or six groups (depending on the therapist variable) was highly significant due to the association of predicted and actual duration. Each contingency table was divided into two smaller tables and comparisons made to test for association of each therapist characteristic with length of therapy among the predicted Quitters and Stayers separately (Cochran & Cox, 1957, pp. 103–105). If results for the two groups differed, an interaction effect of therapist characteristic and patient type on duration would be suggested.

Therapist sex did not relate significantly to duration among predicted Quitters ( $\chi^2 = 2.00$ ,  $df = 1$ ,  $p > .10$ ) or among predicted Stayers ( $\chi^2 = 2.12$ ,  $df = 1$ ,  $p > .10$ ). For the combined patient groups, there appeared to be a relation between therapist sex and duration. However, the  $\chi^2 = 6.40$ ,  $p < .05$ , is not exact as it is based on a comparison made after seeing the data. The results merely suggest that women therapists in this study held higher proportions of patients in therapy.

Profession of the therapist did not relate significantly to duration either among Quitters or Stayers; nor did his experience. There was, however, some suggestion of an experience effect. For the combined patient groups, the association of experience and duration yielded  $\chi^2 = 3.99$ ,  $p < .05$ . Unless confirmed, the as-

TABLE 3  
NUMBER AND PROPORTION OF PREDICTED QUITTERS  
AND STAYERS WHO REMAINED UNDER DIFFERENT  
THERAPIST CONDITIONS

Therapist characteristic	Quitters		Stayers	
	<i>N</i>	<i>p</i>	<i>N</i>	<i>p</i>
Sex				
Male	49	.45	84	.79
Female	8	.62	27	.93
Profession				
Psychiatrist	11	.41	28	.82
Psychologist	28	.44	53	.82
Social Worker	18	.53	31	.82
Experience				
Four years or less	26	.41	46	.81
Over four years	30	.52	67	.87
Personal psychotherapy				
Some	25	.43	59	.80
None	32	.50	54	.81
Competence				
High	11	.46	24	.89
Average	35	.52	48	.75
Low	7	.35	31	.83
Liking				
High	17	.47	51	.88
Low	40	.45	60	.76
Interest				
High	26	.59	63	.88
Low	31	.39	48	.74
"Type"				
A	17	.49	57	.88
B	20	.56	41	.89

sociation also should be regarded as tenuous because it is based on a post hoc comparison and the correlation between experience and length of treatment is low ( $\phi = .12$ ).

Therapists with personal psychotherapy had a median of 13 months of treatment (range = 4–40 months). They did not differ significantly from therapists with no psychotherapy in the extent to which they held either Quitters or Stayers in therapy. Likewise, judged therapeutic competence, therapist "type," and liking for the patient did not relate significantly to duration among Quitters or Stayers. (Chi square = 4.07,  $p < .05$ , for



the association of liking and duration for the combined groups, but, again, the reservation concerning post hoc comparisons applies.)

Remaining in therapy related to the degree of interest the therapist expressed in the type of problem presented by the patient. The relationship was consistently significant both among predicted Quitters ( $\chi^2 = 4.73$ ,  $df = 1$ ,  $p < .05$ ) and among predicted Stayers ( $\chi^2 = 4.14$ ,  $df = 1$ ,  $p < .05$ ). The correlations of interest and outcome among Quitters ( $\phi = .20$ ) and among Stayers ( $\phi = .17$ ) are so similar as to offer little suggestion of an interaction effect.

Some additional analyses were performed to explore for suggestions that any therapist characteristics interact with patient type. All therapists who saw at least two predicted Quitters or Stayers were categorized as obtaining predicted or unpredicted results. A therapist was considered to have achieved a predicted result if he retained at least 50% of his Stayers or lost at least 50% of his Quitters. Two or more predicted Stayers were assigned to 36 therapists. Only 3 therapists failed to retain at least 50% of them. Two or more predicted Quitters were assigned to 29 therapists; 9 of these therapists retained more than 50% of them (actual range = 67%–100%). The proportion (.31) of therapists obtaining unpredicted results with Quitters was significantly greater ( $p < .05$ ) than the proportion (.08) obtaining unpredicted outcomes with Stayers. Stayers appeared to respond as predicted with most therapists in the sample, but there was a distinct group of therapists who somehow retained potential Quitters in therapy.

The above finding does not indicate a significant interaction between therapist characteristics and patient type. It would be necessary to show that Stayers respond differently than Quitters to the nine therapists identified as retainers of Quitters. These nine therapists were assigned 26 Quitters, and only 4 actually terminated. They were also assigned 15 Stayers, 6 of whom terminated prematurely. The proportions of Stayers (.60) and Quitters (.77) remaining with these therapists did not differ significantly ( $.10 > p > .05$ ), although the trend is suggestive.

A final analysis indicated a highly significant

association between the type of patient a therapist most effectively retained and the type of patient assigned to him for treatment. The 16 therapists in the study who retained 100% of their predicted Stayers were initially assigned 47 Stayers and 18 Quitters. The 9 therapists who retained Quitters were assigned 26 Quitters and 15 Stayers. Chi square = 13.21,  $df = 1$ ,  $p < .001$ , for the association between type of patient assigned and type retained. Since, typically, the individual therapist had considerable freedom in rejecting or accepting a specific patient, he was a major factor in the assignment decision. Thus the evidence on patient-therapist compatibility offered no substantial support for the hypotheses that Stayers and Quitters responded differently to the same therapist characteristic. It did, however, uncover two distinct groups of therapists who appeared able to recognize the two patient types, who showed differential preferences in selecting them as patients, and who were unusually successful in retaining the preferred type in therapy.

#### DISCUSSION

In three separate studies, the TR battery has demonstrated useful predictive validity. All three studies have involved veteran populations, and the results may not be generalizable to other populations. Nevertheless, there is reason to believe that the characteristics measured by the TR battery are not unique to veterans and that these characteristics are incompatible with the behavior required in conventional, verbal, insight-oriented psychotherapy interviews. If so, it would appear worthwhile to try the test in other settings as a means of screening out patients for whom ordinary psychotherapy is likely to be simply a waste of time and effort.

While this study revealed that a sizable number of patient characteristics are related to premature termination, few of these added significantly to the predictive power of the TR battery. The one which added most, the therapist's rating of patient motivation for psychotherapy, is not available until the conclusion of 1 therapy hour. Most clinics would prefer to have a prediction in advance



in order to treat the patient appropriately. A sizable proportion of Terminators do not keep their first therapy appointment, and most patients wait 2-4 weeks for their first therapy appointment. It would be preferable to obtain a comparable rating from an intake interviewer, such as the social worker, and to rely on the predictor patterns previously published (Lorr et al., 1958). Addition of the Sociability scale to the predictive battery should await cross-validation, especially since it increases predictive accuracy very little.

The evidence did not confirm Hiler's (1958) report that Quitters and Stayers reacted differently to similar therapist qualities. Therapists who had "more than ordinary" interest in the problems of the patients were more likely to hold each type of patient in treatment. There was suggestive evidence that women therapists, more experienced therapists, and therapists with strong liking for their patients retained higher proportions of both types in treatment. The finding concerning therapist sex seems sample specific. A significant relationship was suggested in the present study, but Lorr et al. (1958) found no such relation, and, in Hiler's sample, women kept proportionately more Quitters and lost proportionately more Stayers. Sample specificity may also account for the experience relationship, as Lorr et al. found no relation between experience and outcome. While the basis for predicting which patients will quit or stay differed from Hiler's, the present method certainly qualifies as a rigorously cross-validated classification scheme based on large samples of cases widely distributed geographically. The liking and interest measures in the present study are not directly comparable to Hiler's peer ratings of therapist warmth. Raush and Bordin (1957), however, have discussed the multidimensionality and unreliability of global assessments of warmth.

The current findings point to an interaction possibility with a distinctly different interpretive significance than Hiler's (1958). There appeared to be subgroups of therapists who somehow selected unusually high proportions of Quitters or Stayers as their therapy patients. They were quite successful in retaining the preferred type of patient in

treatment, whereas the nonpreferred type patient responded to these therapists about as would be predicted from his TR score. Thus the interaction suggested is that different groups of therapists respond differently to the two types of patients rather than the converse. The interaction appears to take place when patients are selected for therapy.

At this time, the sum of the evidence suggests that Remainers and Terminators are two distinct patient populations who generally react differently to what is offered them in the conventional, verbal, psychotherapy setting. Patient characteristics can be assessed which identify large proportions of these two populations. While therapists' reactions to their patients have some influence on the proportions of both populations, they can hold in treatment, there is little evidence that this is a marked effect.

We are at a stage where the potential Terminator can be identified readily and accurately before psychotherapy need be tried as the treatment of choice. Since Terminators constitute a large proportion of patients seeking treatment from clinics and reject verbal psychotherapy as a method of treatment long before it could conceivably modify behavior, it appears there is an urgent need to stop wasting time and effort on a foredoomed treatment procedure. As Bandura (1961) suggests, the time is ripe for exploring and experimenting with other, perhaps drastically different and largely nonverbal, treatment techniques and methods for modifying the symptomatic and problem behavior of these patients.

#### REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- APPLEZWEIG, M. H., DIBNER, A. S., & OSBOURNE, R. L. PEAQ: A measure of psychopathic behavior. *J. Psychol.*, 1958, 14, 26-30.
- BANDURA, A. Psychotherapy as a learning process. *Psychol. Bull.*, 1961, 58, 143-159.
- COCHRAN, W. G., & COX, GERTRUDE M. *Experimental designs*. New York: Wiley, 1957.
- FREEMAN, H. E. Attitudes toward mental illness among relatives of former patients. *Amer. sociol. Rev.*, 1961, 26, 59-66.
- FULKERSON, S. C., & BARRY, J. R. Methodology and research on the prognostic use of psychological tests. *Psychol. Bull.*, 1961, 58, 177-204.

- GUILFORD, J. P., & ZIMMERMAN, W. *The Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply, 1949.
- HILER, E. W. An analysis of patient-therapist compatibility. *J. consult. Psychol.*, 1958, 22, 341-347.
- LORR, M., KATZ, M. M., & RUBINSTEIN, E. A. The prediction of length of stay in psychotherapy. *J. consult. Psychol.*, 1958, 22, 321-327.
- LORR, M., MCNAIR, D. M., MICHAUX, W. W., & RASKIN, A. Frequency of treatment and change in psychotherapy. *J. abnorm. soc. Psychol.*, 1962, 64, 281-292.
- LUBIN, A., & SUMMERFIELD, A. A square root method of selecting a minimum set of variables in multiple regression: II. A worked example. *Psychometrika*, 1951, 16, 425-437.
- MCNAIR, D. M., CALLAHAN, D. M., & LORR, M. Therapist "type" and patient response to psychotherapy. *J. consult. Psychol.*, 1962, 26, 425-429.
- OSBURN, H. G., & LUBIN, A. The use of configural analysis for the evaluation of test scoring methods. *Psychometrika*, 1957, 22, 359-371.
- RAUSH, H. L., & BORDIN, E. S. Warmth in personality development and psychotherapy. *Psychiatry*, 1957, 20, 351-363.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WHITEHORN, J. C., & BETZ, BARBARA J. Further studies of the doctor as a crucial variable in the outcome of treatment with schizophrenic patients. *Amer. J. Psychiat.*, 1960, 117, 215-223.

(Received March 6, 1962)

## A TWO-DIMENSIONAL SEMANTIC DIFFERENTIAL<sup>1</sup>

ROGER B. ALLISON, JR.

*University of Utah College of Medicine*

A "pathology score" summarized the distances of 28 stimulus words from a psychological origin for a modified semantic differential. Several groups of control Ss and psychiatric patients were used to assess various properties of the pathology scores. The scores had internal reliabilities of .77 for control Ss and .88 for psychiatric patients. Test-retest correlations were in the low .60s for control Ss and .45 for patients. Group differences showed a systematic increase in mean pathology scores as the severity of the mental illness increased. The factorial composition was related to clinical diagnosis and inversely to scores on perceptual closure tasks. Further refinement was recommended.

Osgood's theory of the representational mediation processes is intimately connected with association, thinking, and psychotherapy. His semantic differential method also provides an approach to quantifying some of the manifest and latent reactions imbedded in verbal concepts. Both his theory and method (Osgood, 1952, 1953, 1962; Osgood, Suci, & Tannenbaum, 1957) give additional understanding to the clinician who daily witnesses connotative-like responses to core concepts in patients long before the significant features of the concepts become conscious. Although the semantic differential is not capable of answering "why" the individual reacts as he does, it does enable the clinician to make rapid assessments of certain parameters of associations. One disadvantage of Osgood's technique lies in the absence of information about which reactions are frequent and which are atypical. Particularly where individual differences are to be studied, it seems desirable to have a semantic differential in which deviant responses on basic semantic dimensions could be discerned.

This paper presents a modified semantic differential method to be used in investigations of individual differences. The approach is designed to yield information to the following question: If given only two semantic dimensions will the performances of individuals be systematically related to other behaviors which, in part, are predicated upon

the same underlying semantic characteristics? The study was conducted to determine the feasibility of more elaborate projects of less restricted semantic systems, which are now in progress.

### PROCEDURE

The modified semantic differential consisted of a task in which the subject located a series of stimulus words in a two-dimensional semantic space. The good-bad and strong-weak continua were chosen as approximate representatives of the first two semantic dimensions identified by Osgood: *evaluation* and *potency*. The subject made judgments of the following type: Is *stubborn* something you consider to be very bad, bad, neutral, good, or very good? After rating all of the words on the good-bad scale, the subject proceeded to make judgments for the strong-weak continuum.<sup>2</sup> The list of 28 stimulus words appears in Table 1. The ratings required less than 10 minutes for most subjects. The scoring procedures, which are described later, consumed an additional 15 or 20 minutes of clerical time.

The basic similarity between Osgood's method and the present approach is easily recognized. However, rather than assuming that each response scale was an equal-appearing interval scale, each response category for each stimulus word was expressed as a distance from a psychological midpoint which was determined from an adult sample. Likert's (1932) unidimensional scaling procedure was used to convert the percentage of subjects who selected each response option to the average distance from the mean of a normal distribution which corresponded to that particular percentage. The normal deviate values were then translated to a scale having an arbitrary mean of 30 and standard deviation of 10. The scale values, which are listed in Table 1, were obtained from 350 university students who attended a summer school course in introductory psychology. The scale values constituted deviation scores

<sup>1</sup> This research was supported in part by grants from the United States Public Health Service and Scottish Rite Foundation. Special acknowledgment is due N. Dee Lloyd, Edith Jones, and Stephen J. Donoviel, who served as research assistants.

<sup>2</sup> In future studies, "extremely" will be used as a modifier instead of "very" because of its greater scalar effect upon adjectives (see Cliff, 1959).



TABLE 1  
SCALE VALUES OF RESPONSE CATEGORIES

Stimulus word	Very bad	Bad	Neu-tral	Good	Very good	Very weak	Weak	Neu-tral	Strong	Very strong
STUBBORN	12	27	40	50	60	10	23	30	37	49
POLITE	0	3	9	19	35	0	8	16	25	39
TALKATIVE	3	15	26	39	54	6	18	28	38	51
FLASHY	12	24	35	46	57	12	24	34	42	51
EAGER	0	6	15	27	41	0	6	15	28	43
SOFT	6	14	23	33	44	10	20	30	39	49
ELASTIC	3	12	24	36	46	7	15	26	37	48
SMART	0	3	12	22	36	0	3	14	25	38
MINOR	3	14	31	47	57	7	20	34	49	57
EASY	6	16	26	37	49	11	23	33	42	53
COOL	0	13	24	35	48	6	17	28	40	53
REAL	0	0	15	32	41	0	3	16	33	40
BURNING	13	25	36	46	54	9	19	28	36	45
VALUABLE	0	3	15	26	39	3	9	17	28	40
EMOTIONAL	8	20	30	40	52	11	22	29	37	47
SAD	15	29	39	49	60	13	26	35	42	51
THIRSTY	13	26	36	45	54	11	21	31	39	48
CHEAP	20	27	44	57	60	19	28	40	46	53
FAIR	3	12	21	30	40	3	12	21	30	42
CUTE	0	6	18	31	46	0	11	24	35	49
LADY	0	0	14	26	39	6	13	23	32	43
WHITE	0	3	21	33	44	3	12	24	35	45
NIMBLE	0	7	19	32	45	6	15	24	34	47
FANCY	6	13	25	38	53	6	18	29	40	53
SINCERE	0	0	6	16	33	0	6	12	19	34
HAPPY	0	0	3	16	33	3	8	13	22	36
OLD	9	18	30	40	52	12	23	32	42	51
SILENT	7	17	28	40	52	8	18	27	37	48

Note.— $M=30$ ,  $s=10$ .

from a point above and below which 50% of this sample responded. The procedure of allowing the scale values for response categories to vary across stimulus words was one of the major departures from Osgood's method. This modification permitted association disturbances to be inferred from responses which were markedly deviant.

The results of the unidimensional scaling procedures question the appropriateness of the assumption of equal-appearing intervals used by Osgood. The values reported in Table 1 strongly implied that the psychological meaning varied considerably by response category and by stimulus word. For example, "very bad" ranged from 0 for POLITE to 20 for CHEAP, and "very good" ranged from 33 for SINCERE to 60 for STUBBORN. The variability in the mean semantic values reported by Jenkins, Russell, and Suci (1958) may also be submitted as support that scale values for responses be determined separately for each word. However, if the responses are summated over words to obtain a generalized scale for the response categories, the assumption of equal-appearing intervals is justified, as adequately demonstrated by Messick (1957).

For individual use, the two-dimensional semantic

differential was administered by having the subject mark one of the five response categories for the two continua for each stimulus word. His responses were then transcribed onto a scale sheet similar to Table 1 which provided two scores for each word. These two scores served as coordinates to locate the stimulus words on a two-dimensional graph whose axes were the good-bad and strong-weak continua. Figures 1 and 2 represent the results obtained from two psychiatric patients and two control subjects. The centers of these grids correspond

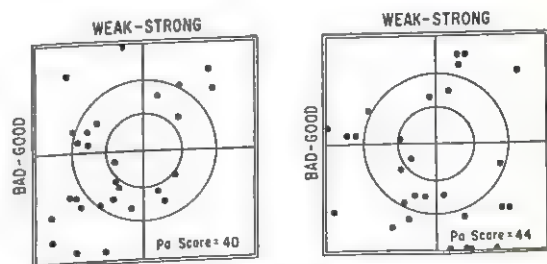


FIG. 1 Scatterplots for two psychiatric patients (paranoid schizophrenic).

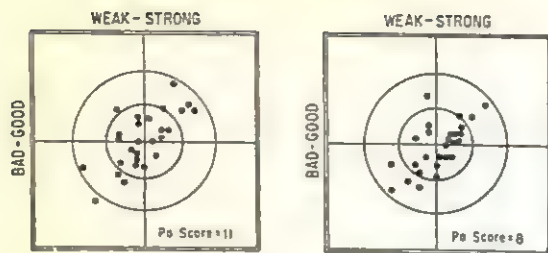


FIG. 2. Scatterplots for two control subjects.

to the joint means of 30 and 30 on the two axes. The inner and outer circles correspond, respectively, to radial distances of one and two standard deviations (or 10 and 20 scaled units) from the center.

In order to reduce the subject's responses to a single summary score, an approximation to the average radial distance of the responses was established: each point was assigned a weight of 0 if it fell within one standard deviation from the joint means, a weight of 1 if between one and two standard deviations, and a weight of 2 if beyond two standard deviations. The sum of the points so weighted was defined as a "pathology score" which, taken at face value, was a measure of the degree to which the meanings deviated from the average. Thus, the higher the pathology score, the greater were the deviations; and the more likely the possibility that an association disturbance existed. Various properties of the two-dimensional semantic differential, particularly the pathology score, are considered in the remainder of this paper.

## RESULTS

### Reliability

Pathology scores based upon the odd items and the even items were determined for 115 university students and 114 psychiatric patients. The students were a different group of subjects than the students used for scaling purposes. The university group had means of 8.8 and 9.5 and standard deviations of 4.0 and 3.5 on the two halves. The correlation between the two sets of scores was .63. Correcting this value for the length of the task resulted in an internal measure of reliability of .77. The values for the patient group were: means, 9.8 and 9.9; standard deviations, 4.7 and 4.3; correlation, .79; corrected correlation, .88.

The difference between the correlations for the university group and the patient group was probably associated with the differences in variability between the two groups. Inspection of the bivariate scatterplots showed all distributions to be skewed toward high

scores although there was less skewness for the patient group.

In general, the odd-even split resulted in nearly equivalent halves with internal reliability adequate for experimental purposes. The higher correlation for the patient group was attributed to the greater heterogeneity exhibited by the patients. Further research may support an argument that the higher correlation resulted from a restriction in semantic freedom for the patients.

Test-retest correlations of the pathology scores were studied with a twofold purpose in mind. One purpose was to estimate the stability of the scores across time and the other was to assess possible changes in pathology scores which were associated with short-term psychiatric hospitalization and treatment. Two groups of university students ( $N = 50, 58$ ) were tested; one group was retested 3 weeks later, the other 8 weeks later. In addition, a group of psychiatric patients ( $N = 47$ ) was tested soon after admission to an inpatient unit and again a day or two before discharge, on the supposition that some clinical improvement had been made. The length of hospitalization for the patients ranged from 3 days to 12 weeks with an average of 2.5 weeks.

The test-retest correlations for the university groups were .61 and .64, which were considerably less than the reliabilities of conventional measures of personality or intelligence. Inasmuch as the number of items was small, the low reliability was not unexpected. If lengthening the task would increase the reliability by the Spearman-Brown formula, five times as many words would have to be used to obtain a reliability of .90.

Although the reliability coefficients for the student groups were low, the coefficient for the patient group was even lower ( $r = .45$ ) and suggested that patients were less stable in their responses than nonpatients.

A comparison of the percentages of subjects who earned lower scores on retesting (60% and 40% for the two student groups as compared with 62% for the patients) offered little evidence that patients improved more over time than subjects who were not involved in a treatment program. Simply using as a measure of improvement the fact that a person was



discharged from an inpatient unit was undoubtedly inappropriate. Some patients were discharged as not improved or transferred elsewhere for further hospitalization and treatment. A more refined approach noting the condition of the patient at the time of discharge should be used in future studies of the relationship between changes in pathology scores and clinical status.

In summary, the reliability of the pathology scores was considered to be adequate for experimental purposes but would need to be increased for operational use. It was difficult to interpret the findings that the internal measure of reliability was higher for patients than nonpatients but the test-retest indices were higher for nonpatients. Further research and understanding of semantic functioning could clarify the differences.

### Group Differences

During the last few years the two-dimensional semantic differential was administered to nine adult groups. These groups are listed in Table 2, along with the respective means and variances of the pathology scores. The groups referred to as Normal Control Subjects and Inpatient Psychotics I were subjects from another study (Cole, Allison, Gortatowski, & Branch, 1960). The assignment of subjects to either group was based upon the results of joint clinical interviews conducted by two senior psychiatrists. The University Students consisted of the two student groups which were used to study test-retest characteristics; the scores which they earned on the first administration of the semantic differ-

ential were used in this analysis. The remaining groups were psychiatric patients whose diagnoses fell within the broad categories listed in the table. Diagnoses were made on the basis of routine clinical procedures of the inpatient and outpatient services. In all cases, the diagnoses were determined independently of any psychological test data reported in this paper.

The data presented in Table 2 showed a consistent trend: as the severity of the mental illness increased there was a corresponding change in semantic functioning as represented by an increase in the mean pathology score.<sup>3</sup>

The data of Table 2 further suggested a relationship between the mean pathology scores and their variances across the diagnostic groups. This is a frequent finding in clinical studies and perhaps meant that as the severity of the mental illness increased an ever-increasing range of test scores was subsumed under the general diagnostic groupings.

The correlated means and variances prompted a transformation of the form  $y =$

<sup>3</sup>The editor has questioned the relationship between this study and Berg's (1955, 1957) *deviation hypothesis*. One similarity between his research, this study, and countless other studies is the significant differences reported between psychiatric patients and nonpsychiatric subjects. When such differences occur they could be construed as supporting the deviation hypothesis. However, psychiatric patients are different by definition. The clinical fields also abound with techniques by which labels may be placed upon people. Further measures developed with the sole purpose to confirm the obvious seem inappropriate. It should be recognized that the behavioral congruencies present in "critical" and "non-critical" areas of functioning originate from the same properties of the individual. It is the attempt to identify and understand these characteristics which constitutes the major difference between Berg's approach and that apparently followed by other investigators. In essence, the deviation hypothesis seems to be only a special case of a basic tenet in psychology—namely, the behavior of an individual in a given situation is a combination of the amount of each attribute he has and the extent to which the situation permits these qualities to emerge. The study of individual differences is therefore concerned with the nature of the combination, the dimensions of human behavior, and the types of situation in which the behaviors may occur. Paradoxical as it may be, it is not the purpose of differential psychology to conclude that people are different.

TABLE 2  
MEANS AND VARIANCES OF PATHOLOGY SCORES

Group	N	M	Variance
Inpatient character disorders	15	14.20	15.23
Normal control subjects	60	15.97	28.63
Outpatient neurotics	44	17.18	38.80
University students	108	18.23	43.42
Inpatient character disorders	49	19.31	34.76
Outpatient neurotics	17	19.65	50.23
Inpatient psychotics	41	21.44	52.94
Inpatient psychotics I	44	22.68	79.98
Inpatient psychotics II	30	23.33	68.71



TABLE 3

PRESENCE OF SIGNIFICANT DIFFERENCES IN MEAN LOG SCORES BETWEEN GROUPS

Group	Group							
	1	2	3	4	5	6	7	8
1. Inpatient character disorders								
2. Normal control subjects	No							
3. Outpatient neurotics	No	No						
4. University students	Yes	No	No					
5. Outpatient character disorders	Yes	Yes	No	No				
6. Inpatient neurotics	Yes	No	No	No	No			
7. Outpatient psychotics	Yes	Yes	Yes	Yes	No	No		
8. Inpatient psychotics I	Yes	Yes	Yes	Yes	No	No	No	
9. Inpatient psychotics II	Yes	Yes	Yes	Yes	No	No	No	No

$\log_{10} x$  before analysis of variance techniques were applied to ascertain whether significant differences existed between the group means. An  $F$  value of 6.10 was significant beyond the .001 level of confidence and indicated that the mean scores between some of the groups were statistically different.

Kramer's extension of Duncan's range test (McGuigan, 1960, p. 185) was employed to determine which means were significantly different at the .05 level of confidence. The results, which are presented in Table 3, indicated that the psychotic subjects had higher scores than normal controls, outpatient neurotics, and university students. However, the attempt to differentiate neurotic subjects from normal people was not successful.

In addition to the quantitative findings, certain relationships were noted between the patterning of responses on the two-dimensional graphs and personality characteristics. Although the relationships must be treated as hypotheses which evolved from the study, they warrant mentioning because of the implications they have for semantic functioning and personality.  $H_1$ : Character disorders reverse the good-bad continuum (e.g., CHEAP is "very good," SMART is "very bad," etc.).  $H_2$ : Depressed people give a preponderance of responses on the bad side of the good-bad continuum.  $H_3$ : Male homosexuals have a negative correlation between the two semantic axes—that is, concepts rated as "good" are also rated as "weak," and concepts rated as "bad" are also rated as "strong."  $H_4$ : Apprehensive individuals mark the "neutral" cate-

gory for most words on the strong-weak continuum. Additional research is required to test these hypotheses.

#### *Correlations with Other Variables*

The semantic differential was included in two other research projects in order to obtain zero-order correlations with as many other variables as possible. The first study utilized a group of 88 adults who were carefully evaluated by psychiatrists and assigned to one of the following categories: Normal, Neurotic, Character Disorder, and Psychotic. Some of these subjects were used in groups reported in Table 2. The pathology scores had significant correlations, at or beyond the .05 level of confidence, with the following variables: Diagnosis (.42), Wechsler Adult Intelligence Scale (WAIS) Comprehension (–.32), WAIS Arithmetic (–.26), WAIS Similarities (–.28), WAIS Verbal IQ (–.28), Gestalt Completion (–.29), Mutilated Words (–.30), and number of literal responses on Benjamin's Proverbs (.32). Nonsignificant correlations were obtained with sex, age, education, WAIS Information, WAIS Digit Span, WAIS Vocabulary, and number of concrete responses to a linear schematization task.

The relationship between the pathology scores and diagnoses represented a correlational statement of the group differences reported in Table 2. The importance of this relationship was the fact that the two-dimensional semantic differential provided an experimentally independent measure of psychopathology.

The other correlations showed that the pathology scores were related to several variables which were also related to diagnosis. For example, the disturbed person generally performs more poorly than other subjects on tasks requiring abstract thinking, judgment, concentration, attention, and the organization of visual percepts. Thus, the pathology scores could have been duplicating information available from other sources. To understand more fully the novel contribution which the pathology score had with respect to diagnosis, a factor analysis of the 15 experimentally independent variables discussed above was performed. The matrix of intercorrelations, with unities in the principal diagonal, was factor analyzed by standard computer methods. The five highest eigenroots, which accounted for 74% of the total variance, and their corresponding vectors were retained and rotated to simple structure. None of the 15 variables had major loadings on more than one rotated factor; the correlations between the rotated factors were  $\leq .31$ .

One of the rotated factors was an expected verbal comprehension ability; the remaining four factors were not defined sufficiently to permit interpretation. Of major interest here was the nature of the loadings for the semantic differential and diagnosis. These two variables defined one of the four factors which were not interpreted and had loadings of .82 and .64, respectively. Two other variables had smaller but negative loadings on this factor: Gestalt Completion ( $-.50$ ) and Mutilated Words ( $-.34$ ). It was quite possible that this factor represented a disturbance in thinking but would require major loadings of additional measures of association, thinking, and perception before it could be interpreted as such. This factor suggested that the semantic differential and diagnosis were measures of the same underlying attribute. Further efforts to clarify this factor are obviously desirable.

The second study was based upon a group of 52 adults who sought psychiatric assistance from an outpatient clinic. The variables available for correlational purposes included the 16 "manifest needs" from the Edwards Personal Preference Schedule and the Thorn-

dike Vocabulary Test. "Aggression" on the Edwards schedule had a correlation of .37 with the pathology scores from the semantic differential. None of the other correlations was significant at the .05 level of confidence. Thus, performance on the semantic differential had little, if any, relationship with stated needs of the individual. The absence of a significant correlation with the Thorndike test was further confirmation that the pathology score was not a measure of verbal ability.

In sum, the correlations which the pathology scores had with other measures suggested that the modified semantic differential was of diagnostic importance and tapped other areas of cognitive functioning such as perceptual closure. Whether the "processes" which underlie a disturbance in semantic functioning are similar to those which disturb thinking and perception requires more extensive research.

#### REFERENCES

- BERG, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, 40, 61-72.
- BERG, I. A. Deviant responses and deviant people: The formulation of the deviation hypothesis. *J. counsel. Psychol.*, 1957, 4, 130-135.
- CLIFF, N. Adverbs as multipliers. *Psychol. Rev.*, 1959, 66, 26-44.
- COLE, NYLA J., ALLISON, R. B., JR., GORTATOWSKI, M. J., & BRANCH, C. H. H. Studies of behavior and metabolism of indole derivatives in schizophrenia. *Amer. J. Psychiat.*, 1960, 117, 393-400.
- JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles of 360 words. *Amer. J. Psychol.*, 1958, 71, 688-699.
- LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol.*, N. Y., 1932, No. 140.
- MCGUIGAN, F. J. *Experimental psychology*. New York: Prentice-Hall, 1960.
- MESSICK, S. J. Metric properties of the semantic differential. *Educ. psychol. Measmt.*, 1957, 17, 200-206.
- OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, 49, 197-237.
- OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
- OSGOOD, C. E. Studies on the generality of affective meaning systems. *Amer. Psychologist*, 1962, 17, 10-28.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

(Received March 15, 1962)



# THEMATIC AND COGNITIVE RESPONSES OF GOOD PREMORBID SCHIZOPHRENICS TO CUES OF NURTURANCE AND REJECTION<sup>1</sup>

KENNETH E. LEBOW<sup>2</sup> AND SEYMOUR EPSTEIN

*University of Massachusetts*

21 good premorbid schizophrenics and 21 matched controls were tested with pictures of mother, father, and peer figures represented in nurturant, ambiguous, and rejecting interaction with a boy hero. The schizophrenics were found to have elevated reaction times to peer pictures following mother pictures (.01 level), obtained relatively low scores of p Nurturance (.05 level), produced flat gradients of p Nurturance as a function of cues varying from rejection to nurturance (.01 level), described mother figures as particularly rejecting (.05 level) and tended (.10 level) to produce their poorest responses to nurturant cues and their best responses to ambiguous scenes. It was suggested that good premorbid deny emotion and prejudice others as rejecting as a defense against emotional involvement. Rather than a specific "censure-cue deficit," schizophrenics exhibit deficit for cues of emotional involvement, in general.

The present study has its roots in two sources. One is a paper on the measurement of approach-avoidance conflict (Epstein & Fenz, 1962), in which it was demonstrated that specially constructed stimulus dimensions in projective techniques provide effective tools for the measurement of conflict, and it was recommended that they be used to explore conflict areas in the behavior disorders. The second is the highly interesting work on schizophrenia reported by Rodnick and Garmezy (1957) in the Nebraska Symposium, in which the value of working with structured cues of theoretical relevance to schizophrenia was indicated. Their conclusion that schizophrenics experience a "censure-cue behavioral deficit" is promising for the conflict dimensional approach of the first source, which uses cognitive deficit along a stimulus dimension as one index of conflict.

Following theoretical conceptions of the etiology of schizophrenia (Arieti, 1959), as well as experimental findings (Rodnick &

Garmezy, 1957), it was assumed that a major source of disturbance for the schizophrenic lies in the mother-child relationship. It was further assumed that the most fundamental characteristic of interpersonal relationships, in general, lies in the realm of acceptance and rejection. Accordingly, a dimension of cues varying from nurturance through rejection in reference to the behavior of a mother toward a young boy was constructed. Partly as a control for the mother figure, and partly for investigation in their own right, parallel dimensions were constructed for father and peer figures.

## METHOD

### *Subjects*

Twenty-one male schizophrenics, judged to be cooperative and to comprehend the task, were tested.<sup>3</sup> The schizophrenics were all "good premorbid" as determined by a total score of 15 or less on the five Phillip's (1953) scales of premorbid adjustment. The decision to test only good premorbid was based on the consideration that they would be less disorganized than poor premorbid

<sup>1</sup> This study is part of a project on the measurement of drive and conflict, of which S. Epstein is principal investigator, and which is being supported by Grant M-1293 from the National Institute of Mental Health, United States Public Health Service. K. Lebow, under the supervision of S. Epstein, conducted the study in partial fulfillment of the requirements for the master's degree at the University of Massachusetts. Both authors participated in scoring the stories.

<sup>2</sup> Now at the University of Nebraska.

<sup>3</sup> We wish to express our appreciation for the cooperation of the administrators and patients of Northampton State Hospital. Walter Simon, chief psychologist at the hospital, and member of the committee for the master's problem, was helpful not only in obtaining patients and otherwise facilitating the research, but also in the suggestions he made throughout the research. While we would like to acknowledge our appreciation to the hospital where the controls were tested, they requested to remain anonymous.

and the results more dependable. Twelve of the schizophrenics were diagnosed as paranoid, three as catatonic, and six as chronic, undifferentiated. All were undergoing drug therapy. Length of hospitalization varied from 3-92 months, with a mean of 40.9 months. Age range was 19-45 years, with a mean of 33.8 and a standard deviation of 7.30. Years of education varied from 6-16, with a mean of 11.4 and a standard deviation of 2.54. Vocabulary level on the Shipley (1940) scale, which was used to assess verbal intelligence, varied from 9-38, with a mean of 25.2 and a standard deviation of 9.02. Occupational background was determined by the United States Bureau of the Census (1950) categories. The sample consisted primarily of unskilled laborers and clerical workers, but contained a considerable number of skilled laborers, and a scattering of individuals in other occupations.

The control group consisted of 21 volunteer patients who were under treatment for physical disorders at an Air Force hospital. Duration of hospitalization varied from 3-21 days, with a mean of 14.5 days. The groups were closely matched on range, mean, and standard deviation for age, education, and vocabulary level. They were also matched for number of subjects in occupational categories.

### Materials

The test materials consisted of four practice pictures followed by nine experimental pictures, all 8 × 10 inches in size. The experimental pictures formed three dimensions, one depicting a mother figure, one a father figure, and one a peer group in interaction with a young boy. Each dimension consisted of a nurturant scene, a neutral scene, and a rejecting scene, presented in that order. For the mother figure the scenes of nurturance and rejection were related to food in order to represent the fundamental nature of the relationship with the mother. For the father figure, scenes were selected to represent the role of the father in providing a guide and model for the boy. Peer scenes related to the role of the group in the socialization process. The experimental pictures are presented in Figure 1, where it can be seen that there is little room for ambiguity in the nurturant and rejecting scenes. The stimuli can be further defined by the mean  $\phi$  Nurturance scores obtained by the control group. In order of nurturance, ambiguity, and rejection, the means are as follows: mother figure: 8.4, 6.9, and 4.0; father figure: 8.6, 6.7, and 4.3; peer figures: 8.5, 5.8, and 3.1. It is apparent that levels of nurturance are different within dimensions and similar between dimensions, as intended.

### Procedure

The schizophrenics were individually tested in a private room in their ward. Subjects were asked if they would agree to participate in a study which dealt with telling stories to pictures; only two subjects refused. A tape recorder which was visible but inconspicuous was described as an aid that would

allow the subject to speak freely without being concerned about the experimenter getting down the material.

Control subjects were tested in the hospital wards, where privacy was obtained by drawing a curtain around the subject's bed, and by selecting only subjects for whom the adjoining beds were vacant. To control for examiner practice effects, the two hospitals were visited on alternate days. To control for and evaluate sequence effects, seven schizophrenics and seven controls were randomly assigned to each of the three following sequences of presentation of the dimensions: MFP (mother-father-peer), FPM (father-peer-mother), and PMF (peer-mother-father).

In order to insure the task being understood, and to reduce practice effects, four practice pictures were presented before the nine experimental pictures. An inquiry was conducted for the first four pictures by asking for a beginning, an ending, and a full description of feelings and action. Following this, the experimenter made no further comment after introducing the pictures. The initial instructions were as follows:

This is a test of creative imagination. I will show you some pictures, one at a time, and I would like you to make up a story about each one. Try to tell a complete story—one that has a beginning, a middle, and an end. If you can, tell what happened before the event, what is happening now, how the characters feel, what they are thinking, and what the outcome will be. Just speak your thoughts as they come to mind. Tell the stories from the viewpoint of the young boy who is the main figure in the pictures. When you are finished with a story, place the picture face down. Do you understand what I want you to do? All right, let's begin.

Each picture was introduced with a comment such as,

Can you tell me a story about this boy and his mother [or "this boy and his father," or "this boy and these other boys"]?

The stories were typed verbatim on 5 × 8 inch cards with an identifying code number so that they could be scored blindly. Before scoring, the cards were shuffled, and the experimenter read a sufficient sample to obtain a frame of reference for an initial sorting. The major score,  $\phi$  Nurturance, varied from strong rejection of the boy hero at the low end of the scoring dimension to strong nurturance toward the hero at the high end. Nurturance was defined as "to express sympathy in action, to be kind and considerate for the feelings of others, to encourage, pity, and console. To aid, protect, defend, or rescue an object" (Murray, 1943, p. 10). Rejection was defined as "any act, thought, or feeling that indicates a lack of concern for the child's welfare." Stories were first scored relative to other stories to the same picture on a five-point scale, with the restriction that each pile had to





FIG. 1. The nine experimental pictures. (The pictures form a  $3 \times 3$  design for nurturance of relationship and figure related to.)

contain at least five stories, except for the two end piles. Two judges independently did the scoring. Where disagreement was within two points, averages were taken; where greater than two points, differences were resolved by discussion. Interscorer

reliability coefficients for the individual pictures varied from .54 to .82, with a mean of .67. These are minimal estimates, as combining pictures should increase reliability, as should averaging the ratings of the judges.

Up to this point, stories were scored relative to a particular picture, so that comparisons between pictures could not be made. In order to establish absolute values across pictures, two typical stories were selected to represent each of the scores on a particular picture. The 90 representative stories (two each for five values for nine pictures) were then sorted relative to each other by two judges along an 11-point scale of equal-appearing intervals. Interscorer reliability for this phase of scoring was very high, as indicated by a coefficient of .97. The two representative stories were then averaged and the original 5-point scales were translated into corresponding values on the 11-point scale. In summary, the final scores were on an 11-point equal-appearing interval scale ranging from extreme rejection through extreme nurturance. The midpoint of the scale, six, represented either an absence of nurturance and rejection or an equal balance between the two.

Stories were also rated on a 5-point scale of "Goodness of Response." As the range was similar from picture to picture, no attempt was made to develop an 11-point scale across pictures. Raters were asked to make a global judgment of the goodness of the stories by taking into account the following four elements:

**Perceptual Integration**—the extent to which the characteristics of the stimulus are accurately noted and integrated into the story with no credit to be given for simply enumerating perceptions. Noting details incompatible with the development of the story is to be considered a poor response.

**Organization and Development**—the degree to which events logically and necessarily follow from each other. A good story should have a beginning and an end as well as a description of ongoing events.

**Creativity and Richness**—the degree to which the story reasonably and effectively departs from mere description of the stimulus. The originality of the plot, the vividness of description, the degree of action and drama, and the extent to which the story captures the reader's interest are favorable considerations.

**Formal Presentation**—the degree to which the story is well presented and is free of grammatical errors, inappropriate comments and laughter, improper use of words, speech disturbances, such as stuttering, and inappropriate pauses. Interscorer reliabilities for Goodness of Response varied from .51 to .71 for the individual pictures, with a mean of .63.

In addition to the two scores already described, the following four measures involving time were investigated: reaction time, response time, rate of pauses, and fluency. In a previous study on the measurement of conflict with a specially constructed thematic apperception test (Fenz & Epstein, in press), distinctive gradients were found for reaction time, rate of pauses, and Goodness of Response as a function of a stimulus dimension relevant to the area of conflict. It was concluded that these measures provide promising indices of anxiety induced cognitive deficit. Reaction time, which was taken from

the tape recording, consisted of the interval in seconds between the introduction of a picture and the beginning of a story. Rate of pauses was defined as the number of silences longer than 2 seconds divided by response time. Response time was defined as the interval in seconds between the beginning and end of a story. Response time was investigated in its own right only because it was available as one of the steps in determining rate of pauses. Fluency was measured by the reciprocal of rate of pauses, and is thus an alternate measure of the same variable.

## RESULTS

The data were analyzed by separate analyses of variance for the dependent variables of *p* Nurturance, Goodness of Response, and the four formal measures. The analysis of variance consisted of  $2 \times 3 \times 3 \times 3$  factorial designs for diagnosis (schizophrenic versus control), relationship (nurturance, ambiguity, or rejection depicted in the stimulus), figure (mother versus father versus peers), and sequence (MFP versus FPM versus PMF).

### Formal Measures

Of the four time measures, only reaction time revealed differences between the schizophrenic and control groups. In Table 1 it

TABLE 1  
ANALYSIS OF VARIANCE OF REACTION TIME

Source	df	MS	F
Between subjects			
Diagnostic group (D)	1	1.06	<1.00
Sequence (Q)	2	135.15	<1.00
D $\times$ Q	2	259.37	<1.00
Subjects (Ss) (within groups)	36	384.96 <sup>a</sup>	13.31***
Within subjects			
Figure (F)	2	77.39	2.31
F $\times$ D	2	52.25	1.56
F $\times$ Q	4	108.94	3.26*
F $\times$ D $\times$ Q	4	164.11	4.91**
F $\times$ Ss	72	33.40 <sup>a</sup>	1.15
Relationship (R)	2	211.17	3.55*
R $\times$ D	2	4.62	<1.00
R $\times$ Q	4	16.86	<1.00
R $\times$ D $\times$ Q	4	60.99	1.02
R $\times$ Ss	72	59.39 <sup>a</sup>	2.05**
F $\times$ R	4	92.20	3.18*
F $\times$ R $\times$ D	4	59.29	2.05
F $\times$ R $\times$ Q	8	34.40	1.18
F $\times$ R $\times$ D $\times$ Q	8	22.13	<1.00
F $\times$ R $\times$ Ss	144	28.92 <sup>a</sup>	

<sup>a</sup> Error term for terms above it to the next error term.  
F  $\times$  R  $\times$  Ss was used as error term for other terms.

\*  $p \leq .05$ .

\*\*  $p \leq .01$ .

\*\*\*  $p \leq .001$ .



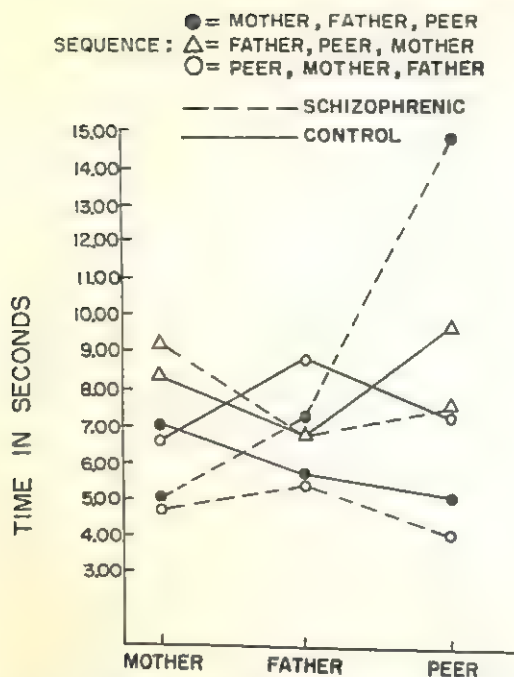


FIG. 2. Reaction time as a function of diagnosis, figure presented, and sequence of presentation.

can be seen that the Diagnosis  $\times$  Figure  $\times$  Sequence interaction is significant at the .01 level. Inspection of Figure 2 indicates that the significance is a result of the highly elevated reaction time of the schizophrenics to the peer figures in the MFP sequence. The results cannot be attributed to fatigue, as neither the mother nor father figures produce similar results when last. The MFP sequence is the only one in which the mother figures precede the peer figures. It thus appears that either the mother cues or the combined mother and father cues have an effect upon the schizophrenics which increases their reaction time to the peer cues.

Other statistically significant findings on the formal measures are of only incidental interest. For reaction time (see (Table 1), relationship is significant at the .05 level because of relatively long reaction times to the ambiguous pictures. It is, of course, not surprising that ambiguous stimuli do not suggest a story so quickly as more highly structured stimuli. A significant Figure  $\times$  Relationship interaction is a consequence of the ambiguous picture of the peer relationship eliciting a higher reaction time than the other ambiguous stimuli. This suggests that the ambiguous picture of the peer group is simply

more ambiguous or difficult than the other ambiguous stimuli. Inspection of the responses supports such a view, as there are several spontaneous comments about it being difficult to tell if the peer group is welcoming or rejecting the hero. The significant Figure  $\times$  Sequence interaction is solely a result of the elevated reaction time of the schizophrenics to the peer figures in the MFP sequence, which has already been discussed. The only other significant finding is on frequency of pauses, where sequence is significant at the .05 level, with the highest rate of pauses occurring in the MFP sequence.

#### Goodness of Response

Following Rodnick and Garnezy's (1957) conclusion that schizophrenics suffer from a censure-cue behavioral deficit, it was anticipated that schizophrenics would obtain lower scores of Goodness of Response on the rejecting than on the ambiguous and nurturant stimuli. The results fail to support expectancy. While the schizophrenics perform significantly more poorly (.001 level) than the controls on all pictures combined, with means, respectively, of 2.40 and 3.33, the differences associated with stimuli grouped either according to rejection, ambiguity, and nurturance, or mother, father, and peer figures are not significant. There is a tendency (.10 level), however, for the schizophrenics to tell their poorest stories to the nurturant stimuli and their best stories to the ambiguous stimuli. Thus, it is the nurturant cues that are more disruptive for the schizophrenic, and the nonemotional ones which he responds to most effectively. The controls do poorest on the ambiguous cues, which is understandable in that ambiguous cues offer the least material for a story.

#### Press Nurturance

While Goodness of Response provided an evaluation of the disruptive effects of cues on cognition, *p* Nurturance provides an approach to the dynamics and response biases of the schizophrenic. Figure 3 presents the results grouped according to cues of nurturance, ambiguity, and rejection. Two observations are immediately apparent. One is that the schizophrenics produce a less steep gradient than the controls; they attribute less

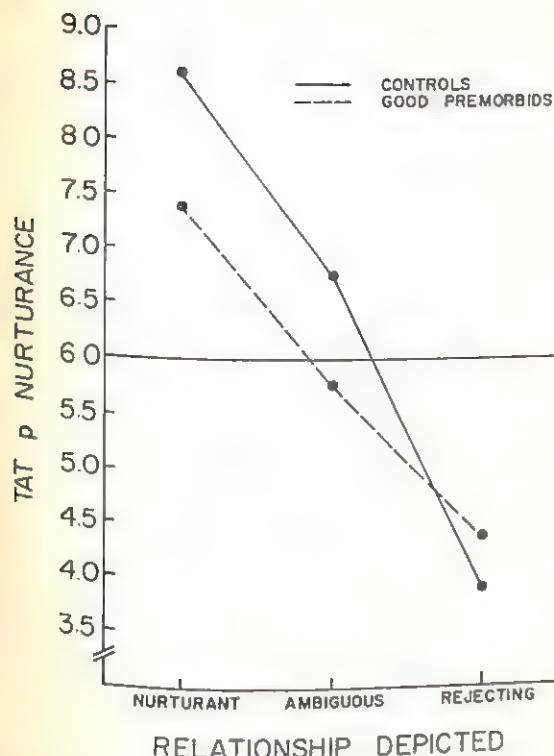


FIG. 3. Press Nurturance as a function of diagnosis and relationship. (A neutral response is represented by a score of 6.)

nurturance to the nurturant stimuli, and less rejection to the rejecting stimuli. The effect is significant at the .01 level, as indicated by the Relationship  $\times$  Diagnosis interaction in Table 2. The other is that the denial of nurturance is considerably greater than the denial of rejection, which corresponds to the finding on Goodness of Response that schizophrenics tend to exhibit greater deficit in responding to cues of nurturance than of rejection. In Figure 3 it can be seen that the schizophrenics obtain lower  $p$  Nurturance scores than the controls on the pooled stimuli. The overall mean  $p$  Nurturance score for the schizophrenics is 5.7 as compared to 6.2 for the controls. Table 2 indicates that the difference is significant at the .05 level. A comparison of the groups on the ambiguous stimuli, which should be free of the effect of avoidance of affect laden cues, reveals that while the controls describe the ambiguous figures as primarily nurturant, the schizophrenics describe them as rejecting (see Figure 3).

A significant Figure  $\times$  Diagnosis interaction (.05 level) in Table 2 indicates that

the schizophrenics and controls differ in the relative nurturance they attribute to the mother, father, and peer figures. The interaction is illustrated in Figure 4, where it can be seen that the groups differ most on mother figures and least on peer figures. The schizophrenics describe the mother figure as highly rejecting, while the controls describe her as highly nurturant.

In order to determine whether the differences between the groups on  $p$  Nurturance are a result of differences in story telling ability, the groups were subdivided into the best 3 and poorest 4 on Goodness of Response within each sequence. The mean Goodness of Response scores for the 12 poorest and 9 best schizophrenics, respectively, are 1.81 and 3.14. The corresponding means for the controls are 2.96 and 3.91. Thus the better schizophrenics outperform the poorer controls. In view of the small and unequal numbers in each subdivision, and the consideration that levels of performance were not orthogonal with respect to diagnostic groups, statistical analysis was not attempted. Rather, the data were inspected to determine whether the relationships already noted held across

TABLE 2  
ANALYSIS OF VARIANCE OF  $p$  NURTURANCE

Source	df	MS	F
Between subjects			
Diagnostic group (D)	1	24.90	4.47*
Sequence (Q)	2	8.94	1.60
D $\times$ Q	2	2.07	<1.00
Subjects (Ss) (within groups)	36	5.57*	3.35***
Within subjects			
Figure (F)	2	6.69	3.98*
F $\times$ D	2	5.93	3.52*
F $\times$ Q	4	2.54	1.51
F $\times$ Q	4	.77	<1.00
F $\times$ D $\times$ Q	8	1.68*	1.01
F $\times$ Ss	72	478.85	146.43***
Relationship (R)	2	23.52	7.19**
R $\times$ D	4	6.34	1.93
R $\times$ Q	4	2.52	<1.00
R $\times$ D $\times$ Q	8	3.27*	1.96***
R $\times$ Ss	72	9.26	5.57***
F $\times$ R	4	1.95	1.17
F $\times$ R $\times$ D	8	3.33	1.88
F $\times$ R $\times$ Q	8	2.80	1.68
F $\times$ R $\times$ D $\times$ Q	16	1.66*	
F $\times$ R $\times$ Ss	144		

\* Error term for terms above it to next error term. F  $\times$  R  $\times$  Ss was used as error term for other error terms.

\*  $p \leq .05$ .

\*\*  $p \leq .01$ .

\*\*\*  $p \leq .001$ .



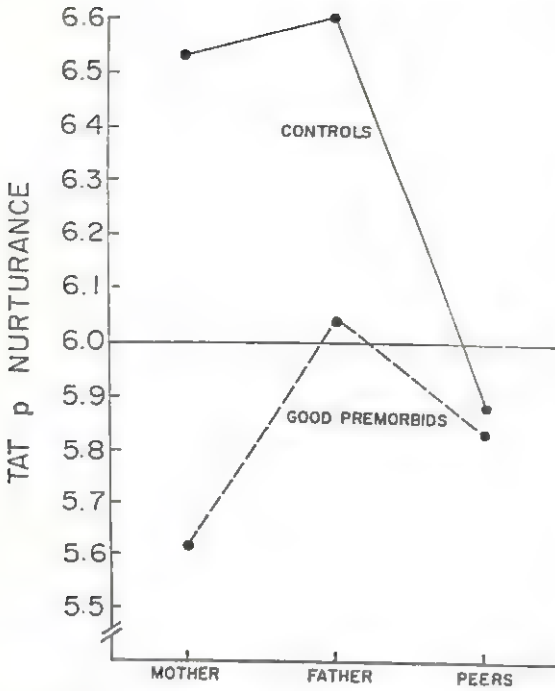


FIG. 4. Press Nurturance as a function of diagnosis and figure. (A neutral response is represented by a score of 6.)

levels of performance. In Figure 5, it can be seen that the differences between schizophrenics and controls on  $p$  Nurturance for mother, father, and peer figures cannot be explained away by differences in story telling

ability, as the best and poorest performing schizophrenics produce parallel curves, attributing least nurturance to the mother figure and most to the father figure. The poor controls, on the other hand, describe the peer figures as least nurturant.

There are two further observations of interest in Figure 5. One is that within both diagnostic groups there is an inverse relationship between adequacy of performance and  $p$  Nurturance. This is not an artifact of scoring bias, as analysis of variance of Goodness of Response failed to reveal a tendency for nurturant pictures to obtain higher scores on Goodness of Response than rejecting pictures. The second observation concerns the distribution of  $p$  Nurturance among the three figures. Whereas the best controls describe the mother figure as the most nurturant, the poorest controls describe her as a relatively neutral figure. Thus, on both overall  $p$  Nurturance and pattern of  $p$  Nurturance, the poorest controls fall between the best controls and the pooled schizophrenics. Of further interest, responses to the mother figure provide the greatest discrimination between the better and poorer performing controls, while responses to the father figure are next, and responses to the peer figures are the least effective discriminators. This is the same

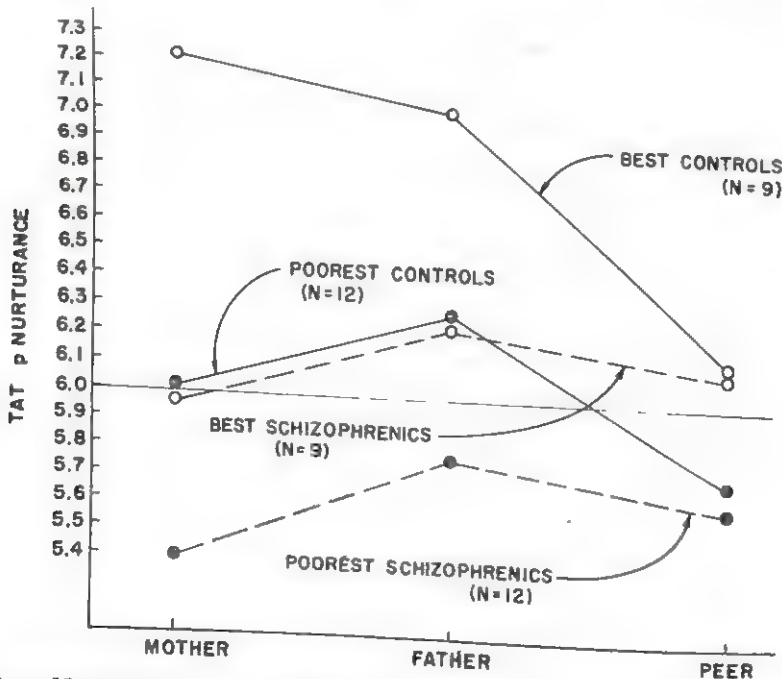


FIG. 5. Press Nurturance as a function of diagnosis and figure, with groups divided on Goodness of Response. (A neutral response is represented by a score of 6.)

pattern as was found for schizophrenics relative to controls.

When  $p$  Nurturance is examined as a function of levels of nurturance in the stimulus, the best controls obtain the steepest and the poorest schizophrenics the least steep gradients. Of more interest, the best schizophrenics, despite higher Goodness of Response Scores, produce a less steep gradient than the poorest controls. Thus, the relatively flat gradient of the schizophrenics cannot be attributed to a general factor of poor responding, but is indicative of an avoidance reaction to cues of nurturance and rejection. Within diagnostic groups, the poorest and best performing subjects differ most on the nurturant stimuli and least on the rejecting stimuli. Thus, again the pattern of  $p$  Nurturance scores of the poorest controls places them between the best controls and the pooled schizophrenics.

There are some significant findings of incidental interest only, in Table 2, which warrant brief reference. As might be expected, the stimulus dimension accounts for a major share of the variance for scores of  $p$  Nurturance. The mean score for the nurturant scenes is 7.90, for the ambiguous scenes, 6.10, and for the rejecting scenes, 4.00. Pooling across levels of nurturance, there are significant differences among the figures (.001 level), with the father figure obtaining the highest score of  $p$  Nurturance, with a mean of 6.20, the mother figure next, with a mean of 5.95, and the peer figures the lowest, with a mean of 5.75. A significant Figure  $\times$  Relationship interaction (.001 level) is a result of a steeper gradient of  $p$  Nurturance for peer figures than for mother and father figures, which may indicate that responses to peer figures are determined by stimulus characteristics to a greater extent than responses to mother and father figures, which presumably are more influenced by personality factors.

## DISCUSSION

### Reaction Time

The elevated reaction time of the schizophrenics to the peer figures in the MFP sequence suggests the possibility of a sequence-cue behavioral deficit in schizophrenia. The MFP sequence was the only one in which the mother figures preceded the peer

figures, so that it appears that either the maternal cues by themselves, or in combination with paternal cues, interfered with responses to the peer cues that followed. It is hypothesized that the sequence-cue behavioral deficit is indicative of a conflict between relating to others and being dependent upon the mother, i.e., thoughts about relating to the mother are incompatible with thoughts about relating to contemporaries. The findings on reaction time are in particular need of verification, as a sequence effect had not been anticipated, and the results are not supported by other measures of cognitive deficit. A study is currently under way in which sequence effects are being more thoroughly investigated in a complete Latin square design, and in which a measure of GSR is used as an alternate index of conflict.

### Press Nurturance

There were three major findings on  $p$  Nurturance. First, the schizophrenics produced flatter gradients of  $p$  Nurturance as a function of nurturant cues than the controls, i.e., they told less rejecting stories to cues of rejection and less nurturant stories to cues of nurturance. Second, the schizophrenics obtained an overall lower score of  $p$  Nurturance than the controls across all figures. While the controls tended to perceive others as nurturant, the schizophrenics described them as rejecting. Third, the schizophrenics described the mother figure as particularly rejecting, while the normals described her as highly nurturant. The first finding is, of course, consistent with the widely held view that the most fundamental defense in schizophrenia is emotional withdrawal. In light of this, the second and third findings are of particular interest, as emotional denial should make the schizophrenic view others as neutral rather than rejecting. Why should the schizophrenic then be prone to describe others, particularly the mother figure, as rejecting? The most obvious explanation is that the schizophrenics were, in fact, rejected by their mothers, and the thematic responses reflect their past experiences. An alternate explanation, not incompatible with the first one, is that denial of nurturance serves as a defense mechanism, which reduces the possibility of personal involvement and rejection. If others



are rejecting there is little reason to attempt to relate to them, and should one make the attempt and be rebuffed, this simply confirms expectation and one is prepared. Thus, a proclivity to perceive people as rejecting, and to deny nurturance, in particular, serves the same function as emotional withdrawal, namely, protection of the schizophrenic from rejection.

Evidence supporting the interpretation of denial of nurturance as a defense mechanism is provided by some recent validity studies on thematic responses (Epstein, 1962; Fenz & Epstein, in press). It was found, for example, that novice parachutists who acknowledged that they were afraid, and were attempting to control their fear, told stories about TAT heroes who were described as unafraid (Fenz & Epstein, in press). The stories, thus, reflected the defense of the parachutist rather than his emotional state. Moreover, as will be seen later, there is considerable evidence apart from the present study that nurturant cues are disruptive to schizophrenics, which is consistent with the assumption that they arouse defensive behavior.

For the mother figure, whom the schizophrenics described as particularly rejecting, denial of nurturance may well be a means of coping with a conflict over dependency. In this respect, it is important to recognize that the schizophrenics are all good premorbid, who would be expected to actively cope with their problems. Such an interpretation is consistent with the findings on reaction time, and is supported by several other studies. In an unpublished study by Ussery, cited in Rodnick and Garnezy (1957), good premorbid schizophrenics told TAT stories characterized by themes of dominance and submission, with heroes attempting to establish independence from mother figures. In a study by Harris (1957) it was found that while poor premorbid overestimated the size of pictures of mother-child relationships, good premorbid underestimated them. Harris interpreted underestimation as indicative of control. In a study on projective responses of paranoid schizophrenics, Lane and Singer (1959) reported, as one of their most clear-cut results, that schizophrenics revealed greater dependence on the mother figure than controls. A

final reason for interpreting low  $p$  Nurturance scores for the mother as a defensive reaction rather than simply as a biographical description is that the schizophrenics attribute considerably higher  $p$  Nurturance to the father than the mother figure, whereas what evidence there is indicates that the fathers of good premorbid are more rejecting than the mothers (Rodnick & Garnezy, 1957). The schizophrenics' thematic responses may thus not only reveal an attempt to cope with dependency by devaluating the mother figure, but may represent an attempt to find a suitable model for masculine identification by favorably evaluating the father figure. It will be interesting to determine whether poor premorbid, who presumably have experienced greater rejection from their mothers than their fathers (Rodnick & Garnezy, 1957), will show a reverse pattern of  $p$  Nurturance for mother and father figures. Presumably, they would be seeking the dependent relationship with a mother figure which the good premorbid is attempting to reject.

#### *Goodness of Response*

The results on Goodness of Response fail to support the hypothesis that schizophrenics exhibit increasing cognitive deficit as a function of decreasing cues of nurturance. Rather, there was a strong tendency for the schizophrenics to tell their poorest stories to the nurturant cues and their best stories to the ambiguous cues. Thus, if anything, the schizophrenics find nurturant as well as rejecting cues disruptive, and the former more so. Rather than a specific censure-cue deficit (Rodnick & Garnezy, 1957), the schizophrenic exhibits a general deficit for cues associated with emotional involvement, whether of a positive or a negative nature. While this conclusion is supported only tenuously by the results on Goodness of Response, there is other evidence in the present study and in other studies which strongly supports such a view. In the present study, the flat gradient of  $p$  Nurturance produced by the schizophrenics is indicative of a failure to vary responses to correspond with changes in the stimulus, and can thus be interpreted as poor responding. In this respect, it is noteworthy that responses to the nurturant cues were poorer

than responses to the rejecting cues. Turning to other studies, Atkinson and Robinson (1961), in an investigation of paired-associate learning, concluded that not only does social reward fail to serve as a positive reinforcer for schizophrenics, but it may actually have a debilitating effect. Turbiner (1961), investigating perceptual discrimination of stimuli varying along dimensions of positive, neutral, and negative affect, found that schizophrenics were poorer than controls on discriminating pairs of stimuli on both negative and positive dimensions, but did not differ on neutral stimuli. In a particularly well-designed study, Johannsen (1961) compared schizophrenics and normals on a double alternation problem under conditions of social and nonsocial feedback. He found that normals performed better with social feedback, and schizophrenics with nonsocial feedback. Thus, the overall picture supports the conclusion that schizophrenics demonstrate deficit for cues of social and emotional involvement in general, rather than only for cues of censure. The reason that some studies find greater deficit for cues of negative affect and others for positive affect may be that the stimuli differ in features other than the affect depicted. In the present study, for example, positive affect for the mother figure was associated with feeding and negative affect with the denial of food. It is possible that if other scenes had been used, nurturant cues would have not been more disruptive and avoided than rejecting cues. There is obviously a need for further work in which cues of nurturance and rejection are presented in a variety of contexts, before it can be established that one set of cues is more disruptive than the other.

Some particularly interesting observations emerged when Goodness of Response was used to subdivide subjects within groups. It was found that the poorest controls fell between the best controls and the pooled schizophrenics in both magnitude and pattern of  $p$  Nurturance scores. While the best performing controls described the mother figure as more nurturant than any of the other figures, the poorest performing controls described her as relatively neutral, and the schizophrenics as highly rejecting. These find-

ings suggest that perception of  $p$  Nurturance, particularly in regard to mother figures, may be related to a very broad dimension of social and intellectual competence, with schizophrenia at one end. It will be interesting to test this hypothesis by obtaining  $p$  Nurturance scores from mental defectives and under-achievers.

## REFERENCES

- ARIETI, S. Schizophrenia: The manifest symptomatology, the psychodynamics and formal mechanisms. In S. Arieti (Ed.), *American handbook of psychiatry*. Vol. 1. New York: Basic Books, 1959. Pp. 455-484.
- ATKINSON, RITA L., & ROBINSON, NANCY M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *J. abnorm. soc. Psychol.*, 1961, 62, 322-326.
- EPSTEIN, S. The measurement of drive and conflict in humans: Theory and experiment. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962*. Lincoln: Univer. Nebraska Press, 1962.
- EPSTEIN, S., & FENZ, W. D. Theory and experiment on the measurement of approach-avoidance conflict. *J. abnorm. soc. Psychol.*, 1962, 64, 97-112.
- FENZ, W. D., & EPSTEIN, S. Measurement of approach-avoidance conflict by a stimulus dimension in a thematic apperception test. *J. Pers.*, in press.
- HARRIS, J. S., JR. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, 1957, 25, 651-671.
- JOHANNSEN, W. J. Responsiveness of chronic schizophrenics and normals to social and nonsocial feedback. *J. abnorm. soc. Psychol.*, 1961, 62, 106-113.
- LANE, R. C., & SINGER, J. L. Familial attitudes in paranoid schizophrenics and normals from two socioeconomic classes. *J. abnorm. soc. Psychol.*, 1959, 59, 328-339.
- MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard Printing Office, 1943.
- PHILLIPS, L. Case history data prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, 117, 515-525.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109-184.
- SHIPLEY, W. S. *Shipley-institute of living scale for measuring intellectual impairment*. Hartford, Conn.: Institute of Living, 1940.
- TURBINER, M. Choice discrimination in schizophrenic and normal subjects for positive, negative, and neutral affective stimuli. *J. consult. Psychol.*, 1961, 25, 92.
- UNITED STATES BUREAU OF THE CENSUS. *1950 census of population: Classified index of occupations and industries*. Washington, D. C.: United States Government Printing Office, 1950.

(Received April 23, 1962)



## TRANSFERENCE IN CO-THERAPY GROUPS

ELIZABETH E. MINTZ

*National Psychological Association for Psychoanalysis, New York*

Literature on co-therapy groups, led jointly by male and female therapists, is discussed. Their advantages, against groups led by 2 like-sexed therapists, include: (a) reproduction of the original family situation, (b) opportunity for patients of both sexes to identify with a like-sexed therapist, and (c) opportunity to work out anxieties with a therapist of the more-dreaded sex. Case histories describe a Don-Juan type of man whose adjustment improved after working through defenses against Oedipal feelings in co-therapy, and an immature borderline woman who overcame panic reactions to both sexes when reassured by the simultaneous presence of 2 parent figures. Special transference phenomena, research possibilities, and possible problems in co-therapy are discussed.

Co-therapy groups, led jointly by a man and a woman therapist, seem logical in view of the wide agreement that a basic value of group therapy is that the patient can experience and work through multiple transferences; indeed, many writers believe also that another basic value of the group is its resemblance to the family. Yet, with some notable exceptions, literature on co-therapy with groups, except for reports on its use in training, usually focuses on its utilization as a kind of parameter in group therapy, a deliberate variation of technique to solve a specific problem or achieve a limited goal.

This is usually done by a prearranged structuring of the situation so that each therapist plays a given role. Thus male co-therapists, one behaving authoritatively and representing the super-ego, the other seeking to represent the ego's integrative function, work together to help institutionalized delinquent boys deal with their ambivalence toward adults (Adler & Berman, 1960). A social worker leaves interpretation to her psychiatrist colleague and serves the patients as a "feeler into the outside world," helping them plan for the future and apparently representing the reality principle (Klapman & Meyer, 1957). A male and female therapist may deliberately seek to resemble the accepted stereotypes of masculinity and femininity. Thus two therapists offer themselves as ideal substitute parents to the mothers and fathers of children under treatment at an agency, and find that questions about feeding and weaning are addressed to

the woman therapist, questions about education to the man (Grunwald & Casella, 1958). Three co-therapists, representing father, mother, and older brother, attempt to parallel the family constellation of the culture in treating hospitalized Porto Rican schizophrenics; the "father" therapist is authoritarian; the "mother" therapist is submissive and offers food; the "brother" is a helpful companion (Maldonado-Sierra, Trent, Fernandez-Maria, Flores-Gallardo, Vigoreaux-Rivera, & De Colon, 1960). The therapist's wife, an assistant rather than a peer, provides nourishment and comfort to the patients in stress situations, appearing as the ideal wife and mother (Fink, 1958).

Besides the special advantages of role-taking, the co-therapy situation may enable one therapist to "supplement and complement the resources of the other" (Demarest & Teicher, 1954); makes it easier for the two therapists to sustain group hostility so intense that it might be difficult for a single therapist to bear (Kassoff, 1958); allows blind spots in one therapist to be counteracted by the other's insight (Loeffler & Weinstein, 1954); and possesses various other advantages as recently enumerated by Mullan and Sangiuliano (1960). Slavson (1960), however, questions the desirability of a "division of the positive and negative transference upon two distinct individuals," believing that a real resolution of ambivalence cannot result from a fragmentation of transference. This viewpoint, in the writer's opinion, might well be opposed on grounds that the "normal"

personality is developed, and ambivalence worked out, in a two-parent situation.

### CO-THERAPY AS A FAMILY SITUATION

Such prestructured co-therapy groups as have been described above, in which the two therapists plan their respective roles in advance, are essentially different from groups which follow the basic psychoanalytic procedure of letting each patient develop transferences according to his own needs and conflicts. Such an approach is reported by Hulse, Lulow, William, Rindsberg, & Epstein, (1956), who found that "double, sexually heterogeneous leadership representing both parental figures seemed to have facilitated the reproduction of the patients' earlier conflicts." Demarest and Teicher (1954) also report that "the presence of male and female therapists made it possible for the patients to structure a family group which allowed the acting out of family conflicts."

Even if both therapists are male,<sup>1</sup> the patients will structure the situation themselves and react differently to the two therapists regardless of whether or not they have decided in advance to play definite roles. One therapist may seem less threatening than the other (Loeffler & Weinstein, 1954); one may seem "good" and one "bad" (Cameron & Stewart, 1955); indeed, some writers state that patients may react to one male therapist as to a father and to another male therapist as to a mother (Lundin & Aronov, 1952). The latter writers find that, though both therapists are male, there is

a simulated family setting created by the presence of two authority figures. . . . The physical characteristics of the therapists become less important than subtle psychological differences. . . . One therapist will be seen as more aggressive and masculine, the other as more protective and feminine (p. 77).

This finding is in line with the general belief that, in individual psychoanalysis, the analyst's sex is seldom crucial; as Fenichel (1945) writes, "Both men and women patients can and do develop both father and mother transferences toward their analysts,

<sup>1</sup> The writer has not found in the literature any report on two women serving as co-therapists.

whether male or female," though he cites some types of patients as exceptions. Mullan and Sangiuliano (1960) also hold that "regardless of whether the therapists are of different sex or not, the patients will conceive of them unconsciously as being of different sex."

Group therapy, however, does not, in this writer's opinion, really duplicate the classical psychoanalytic situation in which the impact of the analyst's real personality is minimized, so that an important part of his function is to serve as a screen for the patient's fantasies and projections. The real personality of the therapist is relatively more important in group treatment. Consequently, many values usually considered characteristic of group therapy are enhanced by the presence of a man and woman as co-therapists, but would presumably be lessened if both therapists are of the same sex.

If group therapy offers an especially good opportunity to work through transference distortions originating in relationships to parents and siblings, the presence of male and female parent figures would seem to be particularly helpful in eliciting and clarifying these distortions. But if the patient reacts, consciously or unconsciously, to one man as a father and to another man as a mother, the unraveling of transference distortions becomes needlessly difficult.

If we believe that the patient can utilize the group as a kind of second family, more accepting and more growth stimulating than his original family, it seems desirable to have both a father and a mother figure present, but a family of like-sexed parents is not acceptable.

And if the group is of value in continually confronting the patient's perceptual distortions with reality, then therapists of opposite sexes offer a unique opportunity for reality testing. But an appreciation of reality can hardly be fostered by a situation in which it is tacitly accepted that, of two male therapists, one will unconsciously be regarded by the patients as a woman.

Indeed, it may well be that confusion about sex identification and sex role, almost always found in people seeking therapy, could well be heightened by a situation in which like-



sexed therapists are seen as playing opposite-sexed roles. But in a group with male and female therapists, both of whom presumably are clear about their sexual identity and reasonably happy with it, a patient of either sex has an especially good chance to introject this sexual self-acceptance.

### CO-THERAPY FOR AN OEDIPAL PROBLEM

The illustrative cases to follow are based on the writer's 3 years of experience with two male co-therapists, comprising three groups in all. This way of working seems a trend among therapists in private practice, usually not because of practical considerations but as a quest for the optimum therapeutic situation. The therapists work on a peer basis, with a group drawn from the private practice of both. Often a patient enters the group after a period of individual treatment, and may continue to see his therapist individually while in group.

The case of "Jack" demonstrates how a patient may use the co-therapy situation to work through a series of transference reactions, which in this case seemed to parallel very clearly the vicissitudes of Jack's relationship to his mother, and also the current difficulties he experienced with women. As with the other case to be described, only those aspects of treatment relevant to co-therapy can be discussed.

Jack, a young businessman, after brief and unsuccessful treatment with a woman, entered treatment with a man, Dr. M<sup>2</sup> who placed him in a co-therapy group. Jack functioned fairly well socially and in business. He sought treatment partly because he experienced anxiety in the presence of male authorities, and partly because he was dissatisfied by his relationships with girls.

Jack's record as a Don Juan was spectacular. He dated several times a week, seldom taking out the same girl more than a few times, and making a determined effort to have sex relations with each girl he dated. If he failed, he lost interest. If he succeeded, which he often did, he soon felt disappointed and again lost interest.

Throughout most of his first year as a patient in the co-therapy group, Jack's attitude toward

Dr. M remained predominantly positive. He idealized and respected Dr. M, adopting the attitude of a boy toward a revered father, though he tended to be petulant and demanding about financial arrangements.

In sharp contrast, Jack's attitude toward Dr. F showed dramatic fluctuations. He began with a markedly flirtatious attitude, a brilliant smile, openly appraising glances. Group members and both therapists all perceived the contempt veiled by Jack's seductiveness. To Jack himself, it was barely below consciousness, and he acknowledged it quite easily when it was pointed out to him. Concurrently, he recognized the contempt and hostility which he felt toward his endless procession of girl friends.

Soon afterwards, Jack began to express open contempt for Dr. F. He found her comments stupid, and did not understand why other group members seemed to take them seriously. He contrasted her ineptitude sharply with Dr. M's skill. "She does things wrong, he does things right." He fantasied that Dr. M was laughing at his female colleague, would be unable to tolerate her inadequacy, and would walk out on the group, taking Jack with him. He recognized this fantasy as deriving from an early childhood wish that his father would get fed up with his despised mother and walk out, taking Jack with him. At this time, Jack was not aware that his wish to be alone with father was essentially a defense against direct Oedipal feelings for his mother, as became clear later in treatment.

Now Jack began to attack Dr. F vigorously for "sloppy sentimentality." He felt she favored her own patients in the group, especially a male patient whom he considered her favorite. This attitude seemed to derive from one of Jack's major childhood problems: a younger sibling, charming but crippled, had apparently been favored by the mother. There was a period of homosexual longing for his brother figure in the group, which Jack was able to acknowledge. He possessed sufficient ego strength to see almost without therapeutic intervention that the homosexual feelings had an unreal quality and were essentially a defense against his wish for exclusive possession of Dr. F's attention. He now recalled that fleeting homosexual impulses, which he had never acted upon, were likely to occur when he was disappointed in obtaining the affection of a woman.

Now, for the first time, Jack began to deal directly with his Oedipal attraction toward his mother. Seated next to Dr. F in the group, he lit her cigarette (contrary to group custom) and was surprised to find himself experiencing some of the fear and excitement he had felt on his first date. He related this to early sexual strivings toward his mother, for which he had feared his father's rebuke, as he now feared Dr. M's. He was aware also that frequently he still felt like a little boy pretending to be a man. Through this experience, Jack realized that he still hesitated to

<sup>2</sup> For convenience, male and female therapists will be called, respectively, Dr. M and F. Thomas E. Tierney was the male therapist in the first case reported; Samuel Shrut in the second case. Grateful acknowledgments are due to both.

compete with male authorities, as he had feared to compete with his father.

Now a new attitude toward Dr. F appeared. At first unwillingly, then openly, he expressed respect. "She does say something good sometimes. . . . I hate to admit it." The contemptuous, flirtatious attitude vanished. About this time, Jack reported a change in his social life. Sexuality for its own sake seemed less important, and he began to spend time regularly with a girl whom he liked and respected. Concurrently, Jack's attitude toward Dr. M became less juvenile, more man to man.

Near the end of his second year of co-therapy, Jack is now beginning to deal with pre-Oedipal problems which had been repressed even more deeply than the Oedipal conflict. He sees that an infantile wish to be dependent, and a concomitant fear of dependency, contributed to his pattern of running from girl to girl to avoid an involvement. He often wishes for Dr. F's sympathy and affection, and recognizes this as an expression of infantile longings. In actuality, he does not try to obtain motherly care from Dr. F, but accepts her as a therapist with friendly appreciation. With occasional setbacks, his business dealings with male authorities have steadily improved, and his relationships with girls have become warmer and more stable.

#### CO-THERAPY FOR A FRIGHTENED PATIENT

The case of "Lois," like Jack's, demonstrates the emergence of a series of transference reactions related to levels of feeling for a significant parent. It also illustrates two special values of co-therapy: it enables a patient to relate to two parent figures simultaneously where this opportunity had been lacking in real life; and it may afford extra security for a patient almost paralyzed by fear.

At 27, Lois came to Dr. F after 3 years of treatment with two male therapists, undertaken because of attacks of agitated depression and panic so severe as to carry her to the very brink of hospitalization. The first therapist had resorted to electroshock and advice. The second, a psychoanalytically trained psychologist, succeeded in helping her to hold jobs intermittently, but finally advised Lois to work with a woman because she was so terrified by the sexuality and hostility of her transference feelings toward him.

Lois had known no men. Her father had deserted the family when Lois was five, and died shortly thereafter. Her only sibling was a sister. Her mother was a simple, immature immigrant woman. The near-psychotic attacks had apparently been precipitated by the marriage of the sister, which had left Lois and her mother alone together, clinging desperately to one another, with almost no other relationships.

Lois dated occasionally, but became panicky and

furiously at any sexual gesture from the man. She could not hold a job for more than a few months, because she grew terrified as soon as she became sexually aware of her boss or any other man in the office.

In 2 years of treatment with Dr. F, there was some improvement. She held jobs longer, seemed further from a psychotic break. But she still found it difficult to endure the presence of men at work or on dates, and she could not break away from the frantic dependency on her mother, now extended to Dr. F also. "Let me come home and live with you . . . I won't make any trouble. I'll do all the housework."

To offer Lois a relationship with a strong male figure in a situation where she could still feel protected by her mother-therapist, Dr. F placed Lois in a co-therapy group. At once, she began to see Dr. M as strong, aloof, unfeeling, and inhuman. Simultaneously, her attitude toward Dr. F began to change. She saw Dr. F, a vigorous woman in the early forties, as pitifully old and frail. Although she knew that Dr. F had a family and a full practice, she saw her as dependent. "When a patient leaves, you just go home and cry and cry. . . . You don't want me to go on a vacation and have fun. You don't want me to get well and leave you. . . . You will die when I leave."

Not until much later in treatment did Lois seem to realize fully that, in fact, Dr. F did not depend on her. Both therapists interpreted to Lois that she was projecting upon Dr. F the same pathological dependency which existed between her mother and herself. They tried to help her see that she herself had contributed to the mutual dependency, and had never been able to admit that her mother could survive without her.

However, insight with Lois did not ever seem as clear or as important therapeutically as it was with Jack. For her, actual experience seemed far more important. Still insisting that Dr. F would die at Lois' departure from treatment, and that her mother could not survive without her, Lois, now almost 30, succeeded with great apprehension in taking an apartment away from her mother. The therapists speculated that this had become possible because, although Lois had blurred Dr. F and her mother into a single image, she also recognized unconsciously that Dr. F was really independent, and was able to use this recognition to correct her distortion of her mother and move out without feeling like a murderess. Indeed, the mother, now in supportive therapy, responded to the separation very favorably.

With Dr. M, Lois, who had been quiet and frozen with her previous male therapists, became gradually more at ease. Though she still saw him as inhuman, she became able to joke and argue with him in the group. She asked for occasional individual sessions with him, granted with the consent of both therapists. And for the first time, she had dates with the same young man for a period of several months.



Perhaps because Dr. M's presence gave security, Lois now began to express frankly a death wish toward Dr. F: "You look tired tonight. Your work is too much for you. Maybe you belong in an old-age home." This death wish reached its peak when she began to talk, only half playfully, of what a pleasure it would be to poison Dr. F "because there isn't room on earth for both of us."

Here co-therapy was terminated by scheduling difficulties, and Lois was offered the choice of working with either therapist. After considerable conflict, Lois chose Dr. M, as both therapists had hoped she would. "I need to learn to get along with men."

A few months thereafter, while working with Dr. M, Lois still sought occasional individual sessions with Dr. F, as if to assure herself that she had not destroyed her therapist mother. In the last of these sessions, she reported gleefully that she no longer found Dr. M aloof and strong. "He's just as weak as you are. Maybe worse. He's only human."

Outside treatment, also, males became human. Lois could now hold a job, because she did not fear her bosses any longer. She began to permit and enjoy physical contact with the man she had been dating, apparently without serious conflict. She developed new interests, traveled on vacations. A year later, Dr. M reports that Lois has finally relinquished her transference distortion, and now sees Dr. F simply as her ex-therapist.

### THEORY AND OBSERVATION

Both these cases illustrate familiar phenomena: the emergence in transference of a series of attitudes toward an important parent figure, each attitude serving as a defense or disguise for a more deeply repressed feeling; concomitant improvement in real-life relationships; lessened anxiety; and the final development of a comfortable, fairly realistic relationship with the therapist. Equally striking cases from co-therapy groups could be described, as well as cases less successful.

But patients improve, or do not improve, in various situations. What special contribution can co-therapy make?

A striking feature of both cases is that co-therapy confronted these patients with a therapist of the sex which they particularly feared. Lois had avoided men, and had made but little progress with two male therapists. Jack had run from genuine relationships with women, and had left his first female therapist. Yet both patients profited by dealing with therapists of the more dreaded sex. Similarly, the writer has found co-

therapy groups especially valuable for certain types of homosexual male patients who have partly worked out their fear of dealing with heterosexual men and are now ready to face a masculine authority figure. Here, the therapist of the less-dreaded sex serves, merely by being present, as a reassurance.

As to the actual vicissitudes of transference, they do not always follow what might have been predicted *a priori*. A series of complementary transference reactions to the two therapists might well have been predicted; perhaps a patient would alternately like and dislike both therapists, or perhaps a patient would feel toward both therapists what he had felt toward father and mother at an earlier age. This latter type of complementary transference does sometimes occur, as exemplified by Jack's fear that Dr. M would rebuke him for relating as an adult man to Dr. F, involving both therapists in a repetition of Oedipal feelings.

Often, however, a patient appears to focus transference reactions predominantly on one therapist at a time. Sometimes a patient uses a positive, stable relationship with one therapist as a background or an anchor in reality, which enables him to experience and express vehement transference feelings toward the other therapist. Thus, for the most part, Jack used his relationship with Dr. M as a background against which he could work out his vivid transference attitudes toward Dr. F. The transference reactions had an intensity and immediacy seldom found in patients whose reality testing is as good as Jack's.

With Lois, co-therapy seemed effective partly because she could relate simultaneously to a father and a mother figure, an opportunity denied her in childhood; and partly because she was able to use each therapist as a shelter or protection against her transference fears of the other. In Dr. F's presence, Lois could learn how to deal with a male authority without panic. And with Dr. M in the group, she could finally experience the paralyzing death wish toward her mother.

Thus, in these patients, the co-therapy situation facilitated two aspects of psychotherapy generally considered effective: with Jack, the development of insight into emotional patterns derived from early

parental relationships, through transference; in Lois' case, a "corrective emotional experience."

Aside from its therapeutic values, the co-therapy group provides a natural laboratory situation which may well contribute to our knowledge of psychotherapeutic processes. The real personality of the therapist, increasingly acknowledged to affect therapeutic events, is easily held constant in co-therapy observations. In private practice, both therapists often conduct groups singly in addition to the co-therapy group. This situation could be duplicated easily in clinics. Thus comparisons could readily be made between Dr. M's solo group, Dr. F's solo group, and their co-therapy group.

In addition to comparisons as to therapeutic gains, specific questions might be answered through recorded sessions. Do patients bring up Oedipal conflicts with more emotional involvement in the co-therapy situation? Do they speak more to the woman than the man in dealing with pre-Oedipal material? Does further observation confirm the writer's finding that transference reactions are expressed with special intensity in co-therapy? If so, does this intensity necessarily imply more therapeutic progress as reflected in behavioral changes?

Finally, it must be noted that co-therapy also presents special problems. For instance, patients might maneuver the two therapists into rivalry, or use a relationship with one therapist as a defense against experiencing feeling for the other. Unresolved tensions between the therapists would heighten such dangers. If co-therapy comes to be used more widely, further study of these aspects of the technique is needed.

# REFERENCES

- ADLER, J., & BERMAN, I. R., Multiple leadership in group treatment of delinquent adolescents. *Int. J. group Psychother.*, 1960, 10, 213-225.
- CAMERON, J. L., & STEWART, R. A. Y. Observations on group psychotherapy with chronic psychoneurotic patients in a mental hospital. *Int. J. group Psychother.*, 1955, 5, 346-360.
- DEMAREST, E., & TEICHER, A. Transference in group therapy: Its use by co-therapists of opposite sexes. *Psychiatry*, 1954, 17, 187-202.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FINK, H. K. Adaptation of the family constellation in group psychotherapy. *Acta psychother.*, 1958, 6, 189-192.
- GRUNWALD, HANNA, & CASELLA, B. Group counseling with parents. *Child Welf.*, 1958, 37, 11-17.
- HULSE, W. C., LULOW, W. V., WILLIAM, V., RINDSBERG, B. K., & EPSTEIN, N. B. Transference reactions in a group of female patients to male and female co-leaders. *Int. J. group Psychother.*, 1956, 6, 430-435.
- KASSOFF, A. Advantages of multiple therapy with adolescents. *Int. J. group Psychother.*, 1958, 8, 70-77.
- KLAPMAN, J. W., & MEYER, RUTH E. The team approach in group psychotherapy. *Dis. nerv. Sys.*, 1957, 18, 95-99.
- LOEFFLER, F., & WEINSTEIN, H. M. The co-therapist method: Special problems and advantages. *Group Psychother.*, 1954, 6, 189-192.
- LUNDIN, W. H. & ARONOV, B. M. Use of co-therapists in group psychotherapy. *J. consult. Psychol.*, 1952, 16, 76-80.
- MALEDONADO-SIERRA, E. D. TRENT, R. D., FERNANDEZ-MARIA, R., FLORES-GALLARDO, A., VIGOREAUX-RIVERA, J., & DE COLON, L. S. Cultural factors in the group-psychotherapeutic process for Puerto Rican schizophrenics. *Int. J. group Psychother.*, 1960, 10, 373-382.
- MULLAN, H., & SANGIULIANO, IRIS. Multiple psychotherapeutic practice: Preliminary report. *Amer. J. Psychol.*, 1960, 14, 550-565.
- SLAYSON, S. R. Discussion. *Int. J. group Psychother.*, 1960, 10, 225-226.

(Received May 9, 1962)



## LEVEL OF ASPIRATION AND SOCIAL DESIRABILITY IN CHRONIC SCHIZOPHRENICS<sup>1</sup>

RUSSELL P. NORMAN

*Queens College, Charlotte, North Carolina*

The relationship between response to social desirability (SD) and level of aspiration behavior (LA) was studied in a chronic schizophrenic population. 107 hospitalized schizophrenics were given a modified form of the Edwards SD scale and were rated as to mental health (MH) status. 4 experimental groups of 12 Ss each were formed representing the high and low extremes on each variable. All Ss were given 10 trials on a letter-digit substitution task and their LA was obtained after each trial. Results indicated that the performances of the high SD Ss were superior ( $p < .05$ ), and discrepancy scores (LA minus performance) were significantly related to the MH variable ( $p < .01$ ), with low MH groups having negative and high MH groups positive discrepancy scores. The use of MH in combination with SD was valuable in demonstrating the existence of several major behavioral variations in a group of chronic schizophrenics.

With the increased use of objective personality inventories, psychologists have become aware of test-taking attitudes or response sets which affect individual performances on these instruments. The response to social desirability (SD) has been the subject of much study in both normal and psychiatric populations. Clinicians have found that in working with psychiatric patients, many appear strikingly normal when responding to the objective type personality inventory (Feldman, 1956; Friedman, 1953; Hillson & Worchel, 1957; Meehl, 1946). Leary (1957) adds emphasis to these findings when he comments that many psychotics, especially those with a paranoid diagnosis, seem to present a two-layer picture; a facade of normality covering intense feelings of rage and despair.

Some rather strong evidence has been gathered linking SD with self-concept, (Cowen & Tongas, 1959; Norman, 1961; Zuckerman & Monashkin, 1957). Subjects who achieve high SD scores appear to have more favorable self-attitudes than subjects with low scores. If it is true that psychotics as

individuals do not all appear to be psychotic on objective psychodiagnostic tests, that is, if their scores are distributed with some high degree of variance, one might postulate certain behavioral correlates of these responses. It should follow that if an individual is suffering from a severe depletion of self-esteem, this would be revealed by means of a task calling for goal-striving and achievement behavior. On the other hand, if the individual were able to erect some defenses against this unfavorable self-concept, he might give quite a different picture on the same tasks.

The level of aspiration technique (LA) has been used frequently in the experimental study of personality. Most investigations have concentrated on the use of this method in terms of achievement motivation. However, it has been suspected by many that LA is an excellent measure of defensive behavior, (Harris, 1955; Holt, 1946; Sears, 1940). Several investigators have used the LA technique in the study of schizophrenia, the more noteworthy studies being those of Jost (1955), Miller (1951), and Winder (1952). The most consistent findings concern the extreme variability of LA behavior in the schizophrenic groups, where LA tended to be unusually high or low in comparison with normal subjects. However, the problem of discovering which schizophrenics set high

<sup>1</sup> This paper is based on a dissertation submitted to the faculty of the University of North Carolina in partial fulfillment of the requirements for the degree of Doctor of Philosophy. For their helpful assistance, the author would like to express his appreciation to W. G. Dahlstrom, E. L. Streuning, Harold Wilensky, and Leonard Solomon.

levels, which ones set low levels, and which ones respond normally is yet to be resolved. It seems that one of the main drawbacks has been the tendency to classify subjects in terms of the standard nosological categories. If it is true that schizophrenics as a group show a great deal of variability on a variety of measures, other methods of selecting and classifying subjects may be more useful in dealing with these particular variables.

It seems that an important dimension to control in research of this type is actual level of mental health. Since schizophrenics show so much variability in degree of illness, it is possible that SD differences will not only reflect varying predispositions toward defensive test responses, but also may be associated with variations in the actual presence of particular symptoms or undesirable behavior. This "reality" factor in social desirability has been mentioned by several investigators (Crowne & Marlowe, 1960; Sperber & Spanner, 1960) and certainly seems worthy of consideration.

The present study was designed to test one rather general proposition; that level of aspiration behavior would be associated with response to social desirability in chronic schizophrenics. More specifically, that high SD subjects would show higher levels of aspiration and positive discrepancy scores (difference between actual performance and LA), while low SD subjects would be more cautious, demonstrating very low or even negative discrepancy scores. It was also felt that these differences would be emphasized by considering the mental health variable. That is, high SD subjects who are relatively healthier will show relatively more expansiveness, and low SD subjects who are relatively less healthy will show relatively more restrictiveness.

## METHOD<sup>2</sup>

A modification of the 39-item Edwards (1957) SD scale was used. The change represented an attempt to create a scale evenly balanced as to responses keyed "True" and "False." Ten False responses were randomly dropped and 11 True responses were substituted from the original 79-item SD scale (Fordyce,

1956). Thus, the final balanced scale contained 40 items.

An independence of functioning scale developed by Wilensky and Solomon (1960) was used to rate the mental health of all subjects. This scale concerns the patient's ability to assume independent responsibility in caring for himself, and ranges from ability to live and work outside the hospital (score of 1) to requiring maximum supervision for daily activities (score of 6). Each subject's final rating was an average of three independent ratings gathered from ward personnel (ward psychologist, psychologist trainee, ward physician, or head nurse). Product-moment reliabilities for the ratings ranged from .73 to .82.

The SD scale was administered to 107 chronic schizophrenics at the Veterans Administration hospital, Montrose, New York. Mean length of hospitalization was approximately 5 years. Those subjects who could not respond relevantly to the scale were eliminated from the study. High and low cutoff points were established for the SD scale and the final mental health ratings were dichotomized. Four experimental groups of 12 subjects each were established, representing the four possible high-low combinations of social desirability and mental health.

A simple letter-digit substitution task was used in the experimental situation. This was patterned after the well-known digit-symbol subtest of the Wechsler-Bellevue scale. The task was explained to the subject and he was given 10 30-second trials, his score (number of substitutions attempted) being reported to him after each trial. Following each trial, the subjects were asked to estimate their performance on the next trial. This estimate, which served as the level of aspiration measure, was given in numerical form. Thus, the experimenter would say, "You did — of them that time. How many are you going to try to do this time?" In general, the subjects had little difficulty understanding the level of aspiration concept. At the conclusion of the task, all were told that they had done very well.

## RESULTS

Table 1 shows mean performance, level of aspiration, and discrepancy scores (LA minus performance) for the four experimental groups. One notices that the high SD groups

TABLE 1  
MEAN PERFORMANCE, LEVEL OF ASPIRATION, AND  
DISCREPANCY SCORES FOR THE FOUR EXPERIMENTAL  
GROUPS OVER 10 LETTER-DIGIT SUBSTITUTION  
TRIALS

Group	Mean performance	Mean LA	Mean discrepancy
I High SD, high MH	23.35	25.82	2.47
II High SD, low MH	21.88	21.02	-.86
III Low SD, high MH	18.78	19.90	1.12
IV Low SD, low MH	17.61	17.29	-.32

<sup>2</sup> Since the original study was more extensive in design, only those methods pertaining to this particular report will be cited.



TABLE 2

ANALYSIS OF VARIANCE OF MEAN DISCREPANCY SCORES  
FOR THE FOUR EXPERIMENTAL GROUPS OVER 10  
LETTER-DIGIT SUBSTITUTION TRIALS

Source	SS	df	MS	F
Social desirability	2.30	1	2.30	.44
Mental health	55.66	1	55.66	10.68**
Interaction	13.66	1	13.66	2.62
Within groups	229.39	44	5.21	
Total	301.01	47		

\*\* $p < .01$ .

tend toward superior performances, but that the discrepancy scores seem to reflect even more striking differences, particularly as related to the mental health variable. The high mental health (MH) groups (I and III) maintained a positive discrepancy score throughout the series of 10 trials, although the Group III subjects seemed more restricted and stayed closer to their actual performances. However, the low MH groups (II and IV) show negative mean discrepancies. This effect was most noticeable in Group II which showed positive discrepancies on the first two trials followed by gradual increase in the negative direction. In other words the low MH subjects, by and large, tended to predict scores that were lower than those they had just achieved.

Two separate double classification analyses of variance were performed on these scores. Table 2 depicts the analysis of discrepancy scores. Although these scores do not seem to be related to the SD variable, the MH dimension is significant beyond the .01 level.

TABLE 3

ANALYSIS OF VARIANCE OF MEAN PERFORMANCE SCORES  
FOR THE FOUR EXPERIMENTAL GROUPS OVER 10  
LETTER-DIGIT SUBSTITUTION TRIALS

Source	SS	df	MS	F
Social desirability	231.44	1	231.44	5.01*
Mental health	11.43	1	11.43	.25
Interaction	7.86	1	7.86	.17
Within groups	2030.51	44	46.15	
Total	2281.24	47		

\* $p < .05$ .

However, in examining the performance scores (Table 3) something of the reverse seems to be true. The high SD groups tend toward superior letter-digit performances ( $p < .05$ ), although there is no discernable relationship between performance and MH.

### DISCUSSION

To initiate discussion, it would seem most helpful to examine each of the four experimental groups in terms of the main independent variables and actual behavioral observations.

#### Group I: High SD and High MH

The responses of these subjects approximated those obtained from normal subjects in other studies. The LA was always positive, with discrepancy scores gradually decreasing as performance neared an asymptotic level. These subjects were reserved but cooperative, and it did not seem as if they could be appropriately called "symptom deniers." They appeared to be individuals who had made an excellent adjustment to the hospital community, and symptoms, which are a primary referent for the nonhospital community, may not have been crucial. These seemed to be chronic hospital citizens who had become adjusted to a simple and regimented existence.

#### Group II: High SD and Low MH

In this group, LA behavior was strikingly inconsistent with the picture of symptom denial presented on the SD scale. Many of these subjects showed signs of intense anxiety, and some varied from extremes of near muteness to almost manic speech. Although one must interpret their high self-concept (SD) as a rather shaky and defensive one, their ability to approximate normal test responses allows them to function more adequately and with fewer restrictions within the hospital environment.

#### Group III: Low SD and High MH

These subjects were similar to those of Group I, but their general LA, although always positive, was never as high as that of Group I and their predicted improvements were uniformly small. Miller (1951) found this reaction in his group of paranoid schizo-

phrenic subjects. It appeared to the writer that these individuals demonstrated firmly entrenched negative self-concepts. Their incidental verbalizations contained numerous self-effacing remarks, and although they were cooperative, they felt compelled to constantly remind the examiner of their shortcomings.

#### *Group IV: Low SD and Low MH*

This group was characterized by very small discrepancy scores which changed little over the series of 10 trials. Behaviorally, these subjects appeared more disturbed than those in the other three groups. Many were nearly mute and had difficulty communicating with the examiner. They seemed easily threatened and aware of their illness, and were upset by close interpersonal contact with a stranger. Their aspirations varied little from what they had achieved on the previous trial. Although three of the original subjects in this group seemed unable to understand the task and were eventually replaced, the writer was convinced that the others, although obviously disturbed, were responding relevantly to the experimental situation. In addition, they had already demonstrated their ability to respond meaningfully to a personality questionnaire.

A number of interesting questions seem to be raised as a result of these findings. It seems quite clear that the degree of illness in the low MH groups is reflected in their cautious approach to the experimental task. Regardless of actual performance (on which Groups II and IV did differ) few of these subjects were willing to anticipate any improvement in their scores. In fact, the tendency, especially in Group II, was to become more pessimistic as the trials progressed even though the general trend of performance was one of improvement. As has already been pointed out, this negative discrepancy score is even more interesting in light of the picture of symptom denial (high SD) and relatively superior performance. It is possible that the LA task probed a much deeper emotional level, whereby a "true" self-concept was revealed in contrast to the rather grandiose picture presented on the SD scale. Although it is possible to interpret the Group II pattern as one of symptom denial, lack of insight, etc., there is little doubt that whatever the process might

have been, it has at least allowed these individuals to be more effective within the hospital setting. In some ways they seemed like Group I patients in that they often engaged in productive activity and were able to move about with fewer restrictions.

Another way of interpreting the data would be that the low SD subjects were performing in tune with their own expectations, in that their ability at the letter-digit task was not equal to that of the high SD subjects. With no defenses to bolster their sometimes obvious lack of self-esteem, these individuals may have been relating to the experimental situation in the one way that would have been consistent with their self-image. Undoubtedly, a major question remains as to the possibility of specifying more directly the relationships between SD and self-concept. If low SD is equated with a negative self-concept, why do Group III subjects react more like Group I subjects in their LA behavior, but actually perform less adequately on the task? More research is clearly needed to untangle a possible complex interaction between these factors.

The results of this study would seem to have implications for psychodiagnostic testing as well as for a better understanding of the symptomatology of schizophrenia. While some would interpret these findings as indicating the doubtful validity of objective testing procedures in dealing with this type of patient, the writer feels that an individual's ability to respond as a normal person in the presence of a severe psychosis is certainly a valuable piece of information. In many cases it may be better to interpret the apparent normality of a test response as an achievement, rather than a denial or an indication of some imperfection in the test itself. It would seem, however, that in many cases this information could not be properly evaluated in the absence of additional data gathered through the use of more penetrating techniques. The SD-LA discrepancy in Group II was a clear demonstration of this fact.

There is a remarkable similarity between these findings and those of other investigators in the area of schizophrenic symptomatology (Guertin, 1955; Guertin & Zilaitis, 1953; Kleinmuntz, 1960). Since traditional nosolog-



ical categories fail to take normality factors into account, an SD-MH comparison or some other such breakdown might be more useful. Guertin's four categories of schizophrenic behavior (socially normal, grandiose and deluded, evasive well-integrated, sensitive, inadequate and withdrawn) are highly similar to those categories developed in the present study. Although, the former were derived from MMPI responses by factor analytic methods.

## REFERENCES

- COWEN, E. L., & TONGAS, P. The social desirability of trait descriptive terms: Applications to a self-concept inventory. *J. consult. Psychol.*, 1959, 23, 361-365.
- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, 24, 349-354.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- FELDMAN, M. J. The use of the MMPI profile for prognosis and evaluation of shock therapy. In G. S. Welsh & W. G. Dahlstrom (Ed.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, 20, 171-175.
- FRIEDMAN, I. Phenomenal, ideal and projected conceptions of self, and their interrelationships in normal, neurotic, and paranoid schizophrenic subjects. Unpublished doctoral dissertation, Western Reserve University, 1953.
- GUERTIN, W. H. A transposed analysis of the Bender-Gestalt performances of paranoid schizophrenics. *J. clin. Psychol.*, 1955, 11, 73-76.
- GUERTIN, W. H., & ZILAITIS, V. A transposed factor analysis of paranoid schizophrenics. *J. consult. Psychol.*, 1953, 17, 455-458.
- HARRIS, M. W. Protective mechanisms utilized in reactions to ego-threatening situations, as evidenced by performance on a level of aspiration problem. *Dissert. Abstr.*, 1955, 15, 1116-1117.
- HILLSON, J. S., & WORCHEL, P. Self-concept and defensive behavior in the maladjusted. *J. consult. Psychol.*, 1957, 21, 83-88.
- HOLT, R. R. Level of aspiration: Ambition or defense? *J. exp. Psychol.*, 1946, 36, 398-416.
- JOST, K. C. The level of aspiration of schizophrenics and normal subjects. *J. abnorm. soc. Psychol.*, 1955, 50, 315-320.
- KLEINMUNTZ, B. Two types of paranoid schizophrenia. *J. clin. Psychol.*, 1960, 15, 310-312.
- LEARY, T. *The interpersonal diagnosis of personality*. New York: Ronald, 1957.
- MEEHL, P. E. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.*, 1946, 30, 517-524.
- MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, 46, 378-387.
- NORMAN, R. P. Level of aspiration in chronic schizophrenics as a function of social desirability and threat of failure. Unpublished doctoral dissertation, University of North Carolina, 1961.
- SEARS, PAULINE S. Level of aspiration in academically successful and unsuccessful children. *J. abnorm. soc. Psychol.*, 1940, 35, 498-536.
- SPERBER, Z., & SPANNER, M. Social desirability, psychopathology, and item endorsement. Paper read at Eastern Psychological Association, New York, April 1960.
- WILENSKY, H., & SOLOMON, L. Characteristics of untestable chronic schizophrenics. *J. abnorm. soc. Psychol.*, 1960, 61, 155-158.
- WINDER, C. L. On the personality structures of schizophrenics. *J. abnorm. soc. Psychol.*, 1952, 47, 86-100.
- ZUCKERMAN, M., & MONASHKIN, I. Self-acceptance and psychopathology. *J. consult. Psychol.*, 1957, 21, 145-148.

(Received May 17, 1962)

## BRAIN DAMAGE DIAGNOSIS WITH THE MMG

THOMAS R. SCOTT, ROBERT A. BRAGG,<sup>1</sup> AND ROY G. SMARR

*Veterans Administration Hospital, Columbia, South Carolina*

A method for measuring the speed of the spiral aftereffect is described. Results of measurements on a group of 51 brain damaged and 37 control Ss are presented. Both binocular and interocular transfer conditions were used. It was found that brain damaged Ss varied over a much greater range than did controls. Also, the brain damaged S's aftereffect varied more from time to time. Among controls, interocular transfer aftereffect was found to be as great, if not greater, than the binocular aftereffect. Among brain damaged Ss, the interocular transfer and the binocular aftereffect displayed a patterning which was significantly different from that of controls. Measurement of aftereffect rate by the method described (Metamegethograph) shows promise as an indicator of presence or absence of brain damage.

Wohlgemuth (1911), in writing on visual aftereffects of motion, said ". . . it is evident that an abnormal increase or decrease in the intensity of the aftereffect may afford some valuable aid in (brain damage) diagnosis" (p. 115). This suggestion was not investigated directly until Freeman and Josey (1949) found greater memory impairment in patients who failed to report the aftereffect. Later Price and Deabler (1955) found that patients with brain damage report the spiral aftereffect significantly less frequently than control subjects. Their method of determining presence or absence of the spiral aftereffect has been named the Spiral Aftereffect Test (SAET). In this method the subject views a rotating spiral (the eliciting stimulus). The spiral then stops suddenly and the subject reports the aftereffect as an apparent expansion or contraction of the now stationary spiral (the test stimulus). This work was replicated by Page, Rakita, Kaplan, and Smith (1957); Gallese (1956); Davids, Goldenberg, and Laufer (1957); and by Gilberstadt, Schein, and Rosen (1958). These replications resulted in considerable shrinkage in validity, although the consensus remains that the SAET has some value as a screening device. Other studies relating presence or absence of the aftereffect to brain damage are reported by Standlee (1953) and Goldberg and Smith (1958). These studies will be referred to as the SAET studies.

<sup>1</sup> Now at Veterans Administration Hospital, Salisbury, North Carolina.

While the SAET is brief and easy to administer, it does suffer from certain disadvantages. First, the test result is dependent on the ability of the patient to put his subjective experience into words. London and Bryan (1958) and Holland and Beech (1958) found that by using more specific inquiries they could obtain reports of the aftereffect from virtually all subjects whether brain damaged or not. These findings suggested that the differences between brain damaged patients and normals may have resulted from differences in the ability to organize verbal responses rather than differences in the aftereffect itself.

A second disadvantage is that the method yields numerous false negatives (Gilberstadt et al., 1958; Stilson, Gynther, & Gertz, 1957). An untested but possible explanation for the false negatives is that some brain damaged patients may have abnormally intense aftereffect. Since the SAET is based upon reported presence or absence of the aftereffect, the method identifies as abnormal only those who fail to report it. Using the method of Price and Deabler, patients with abnormally great aftereffect would be grouped with normals.

A third difficulty in this area has been a lack of standardization of test conditions in different laboratories. Indeed, there has existed no easy method of deciding upon which combination of test conditions might be most sensitive to brain impairment. Such parameters as type of spiral, illumination, speed of rotation, fixation time, texture,



adaptation, and distance of the subject from the spiral have rarely been constant between any two experiments. Several of these parameters have been shown to influence the aftereffect significantly (Cann, 1961; Durup, 1928; Scott, 1962; Scott & Medlin, 1960, 1962; Singer, 1959; Stern, 1959; Wohlge-muth, 1911).

Persistence time of the illusion has also been studied in connection with brain damage. This measure has been used by Eysenck, Holland, and Trouton (1957), Spivack and Levine (1957, 1959), and Truss and Allen (1959). These studies will be referred to as the persistence time studies. The first of these did not include brain damaged subjects, but both studies by Spivack and Levine and the study by Truss and Allen found significantly longer durations in the brain damaged subjects, suggesting that brain damaged subjects have *more* aftereffect than normals. The SAET results, on the other hand, suggest that brain damaged subjects have *less* than the average amount of aftereffect. The persistence time studies, then, would at first glance seem to contradict the SAET studies under the apparently reasonable assumption that an aftereffect which is stronger will also persist longer.

Perhaps the difficulties and apparent contradictions mentioned above could be clarified by the perfection of a technique for measuring the speed of the aftereffect in a quantitative manner. Several methods for measuring speed are suggested in the non-clinical literature. Durup (1928) used a subjective rating on a four-point scale for judging "intensity" of the aftereffect. Wohlge-muth (1911) had the subject make comparisons of recently experienced aftereffects with actual motions of visual stimuli presented after the aftereffect had completely dissipated. Wohlge-muth also had the subject make an adjustment of the test stimulus for measuring rotational aftereffect. He did not make extensive use of the method, however. This last method has been extended by Singer (1959) and Cann (1961) for measuring spiral aftereffect using an adjustable iris diaphragm as a test stimulus. Scott and Medlin (1962) have developed a more refined method

of measurement utilizing an electronic test stimulus and a psychophysical method.

The purpose of the present study is to measure the speed of the spiral aftereffect in an objective, quantitative manner which is relatively independent of verbal ability and which is sensitive to abnormally fast as well as abnormally slow aftereffect. Previously obtained psychophysical measurements were used in choosing the conditions of the experiment so as to maximize the aftereffect. Speed of the aftereffect is operationally defined as that rate of change of size of the test stimulus which exactly cancels the aftereffect, thus appearing subjectively of constant size.

## METHOD

### Apparatus

A technique for making psychophysical measurements of the rate of change of perceived size in the spiral aftereffect has been briefly described by Scott and Medlin (1962). With the exception of a few minor modifications, the same method was used in the present study. The apparatus, called the Metamegethograph (MMG), consists of a mirror tachistoscope which alternately presents a spiral and a test stimulus. In this study the spiral was presented for 10 seconds and the test stimulus for 2 seconds. These two stimuli were recycled continuously with no delay between presentations. The spiral used in this study was a four-turn, single Archimedes spiral, drawn counterclockwise outward with black India ink on a white 8-inch disk. It was illuminated by two NE-34 neon bulbs operating at 150 volts dc. The spiral was rotated counterclockwise at 160 rpm. The test stimulus was a circular figure produced on a cathode ray tube (5CP1) by a special signal generator.<sup>2</sup> The circular test stimulus always appeared with an initial diameter of 3 inches and then expanded or contracted uniformly and linearly at any one of 21 different precalibrated rates. The rates used were calibrated in percentage per second units based on the initial circle diameter of 3 inches. The 21 settings ranged (in successive 2% steps) from 20% per second contraction through 20% per second expansion, including a middle setting for no change in circle size.

The rate of expansion or contraction of the test stimulus was selected by a stepping relay having 21 positions, corresponding to the 21 settings just described. The subject could step the relay in either direction, one step at a time, by pressing either of

<sup>2</sup> A 3-page description and one circuit diagram have been deposited with University Microfilms, Inc., Ann Arbor, Michigan. Order Special Film S-294, remitting \$2.75 for microfilm or positive copies.

two buttons. He could thus adjust the rate of change of size of the test stimulus by fixed, equal amounts. An automatic graphic record was kept of the successive positions of the stepping relay throughout the experiment. The push buttons were connected to the stepping relay so that pressing the right-hand button stepped the relay in the direction which, on the next cycle, would produce a circle which was objectively contracting more (or expanding less) than on the preceding trial. The left-hand button had just the opposite effect.

### Procedure

Upon entering the testing room the patient was seated in front of the tachistoscope so that his eyes were 5 feet from the spiral. The manipulation box containing the two push buttons was placed in the subject's lap. With the spiral motor turned off, he was given the following instructions:

I want you to look at the center of this spiral. Do not let your eyes wander away from the exact center. In a few seconds the spiral will disappear and a circle will appear. Sometimes the circle will be expanding (or getting larger), and sometimes it will be contracting (or getting smaller). Each time the circle comes on I want you to tell me whether it seems to you to be getting larger or getting smaller. If the circle appears to be getting larger, press the right-hand button. If it appears to be getting smaller, press the left-hand button. If you are not sure, guess.

The room was then darkened, and with the spiral still stationary a series of discrimination trials was run to make sure that each subject understood the instructions. No subject was included in the study who could not correctly identify 4% expansion or contraction of the circle. Most subjects were able to discriminate 2% per second expansion or contraction. If the subject appeared confused his task was explained again. Great care was taken to make sure that the subject understood the task and was able to respond correctly. Following this training procedure the spiral motor was started and the actual test was begun.

On the first trial the relay was always set to its middle position (the setting producing zero change in circle size). Thus, for example, if the subject has an aftereffect on the first trial which makes the stationary circle seem to expand, he presses the right-hand button. This response steps the relay to the position which will, on the next trial, produce 2% per second contraction. If on the second trial the subject's aftereffect exceeds 2% per second, he will again press the right-hand button, signaling that the circle still appears to expand. This will result in moving the stepping relay to the position which will produce 4% per second contraction. If, for example, the subject's aftereffect is less than 4% per second, the circle will now appear to shrink and the subject will

press the left-hand button. This response will return the stepping relay to the 2% per second contract setting. The subject continues in this manner, thus constantly readjusting the rate of change of size so that it moves back and forth around a value at which his aftereffect exactly cancels the objective motion of the test stimulus. If the objective shrinkage of the test stimulus is thus canceled by the subject's aftereffect rate he will experience subjective size constancy. In this case the subject is instructed to guess. It can be seen that this procedure produces a sort of tracking behavior which bears some similarity to the classical method of limits.

Subjects were required to verbalize their responses as well as to depress the buttons, and the experimenter kept careful check to be sure that the subject's verbal responses agreed with his button responses. The subject was allowed to make 75 consecutive responses under each of three conditions. The binocular condition was one in which the subject viewed both the spiral and the test stimulus with both eyes. There were two interocular transfer conditions during which a shutter was placed in front of the subject so that the spiral was viewed by one eye and the test stimulus by the other. The interocular condition in which the spiral was viewed by the left eye will be called the left-right transfer condition, the opposite case being the right-left transfer condition. The order in which these three conditions were presented to each subject was varied systematically so that approximately equal numbers of subjects were run under each of the six possible orders.

The particular values for the number of turns of the spiral, its speed and direction of rotation, the number of trials, and the timing of the exposure period and test period were chosen on the basis of psychophysical data which had been previously obtained (Scott, 1962).

### Subjects

In all, 107 hospitalized male veterans were tested by the procedure outlined above. These subjects were selected on the basis of admitting symptoms which are likely to be found in patients with brain syndromes or in patients with peripheral neuropathy. Of these 107 subjects, 51 were chosen for the brain damaged group. Before classifying a patient in this group, the diagnostic criteria were critically examined. When any reasonable question existed as to the presence of brain damage the patients were not included in this group. A control group of 37 patients was selected. The diagnostic data on the control subjects was also carefully examined to eliminate doubtful cases in which brain damage did not seem to be thoroughly ruled out on the basis of neurological and laboratory findings. Nineteen patients could not be satisfactorily classified by the above criteria and consequently were not included in the study. Twelve of these were alcoholics in whom presence or absence of brain damage could not be definitely established.



## RESULTS

As mentioned above, a graphic record was kept of the successive positions of the stepping relay. Each position of the stepping relay represented a particular rate of change of circle size. In the subsequent discussion an average of the 75 successive rates obtained from a subject under a given condition will be used as a measure of that subject's aftereffect rate under that condition. Positive values indicate aftereffects of perceived expansion. Negative values indicate perceived contraction.

Figure 1 shows the frequency distributions of aftereffect rate for controls (black bars) and brain damaged subjects (open bars). The results for the three conditions for each subject were included in preparing these distributions. Thus, the number of measures represented in the control group distribution is  $3 \times 37 = 111$ , and the number of measures represented in the brain damaged group distribution is  $3 \times 51 = 153$ . Figure 1 shows that the aftereffect rates for the control subjects had a comparatively narrow range, while the range of aftereffect rates for brain damaged subjects was much greater and extended both above and below the control group distribution. The modal aftereffect rate for brain damaged subjects was somewhat greater than for controls.

The spiral and direction of rotation used would be expected to elicit aftereffect of perceived expansion (indicated by positive aftereffect rates in Figure 1). On a number of occasions the brain damaged subjects evidently perceived aftereffects of contraction (designated by negative rates in Figure 1). These results suggest that some of the brain damaged subjects occasionally experienced aftereffect in the *same* direction as the eliciting motion. Aftereffect of this kind has been reported by Szily (1905, 1907), Wohlge-muth (1911), and Fuchs (1928), although there has been some controversy on this point (Stern, 1959).

Figure 2 compares the aftereffect rate for the two groups broken down also according to the three conditions: the left-right transfer condition (L), the binocular condition (B), and the right-left transfer condition (R). The white bars indicate a range of  $\pm 1$  standard error. As was evident in Figure 1, the variance is heterogeneous, and a conventional analysis of variance would not be appropriate. It is apparent that the average aftereffect rate for the brain damaged patients was somewhat higher than for the controls, but the meaning of this difference is not clear in light of the extremely high variability in the brain damaged group. In the control group, slightly more aftereffect was produced by the transfer conditions

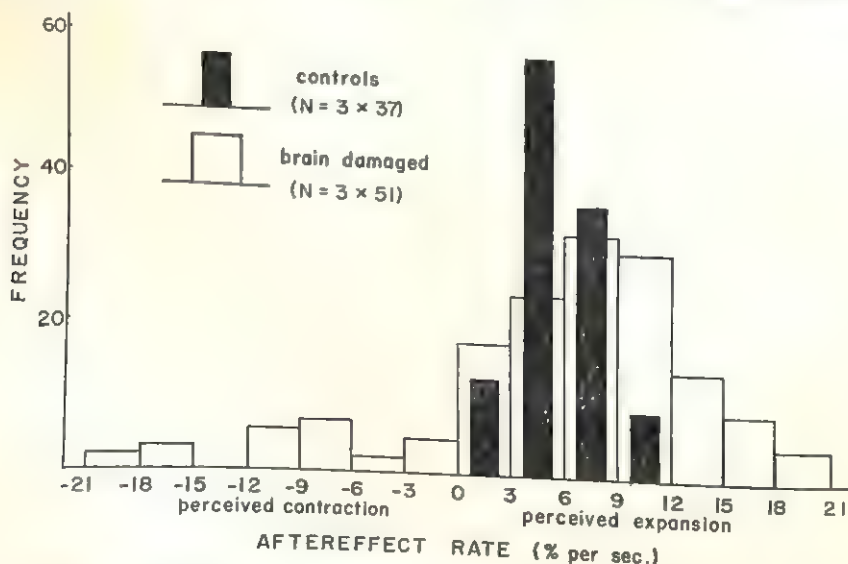


FIG. 1. Frequency distribution of aftereffect rate based on three measures obtained on each subject in both groups.

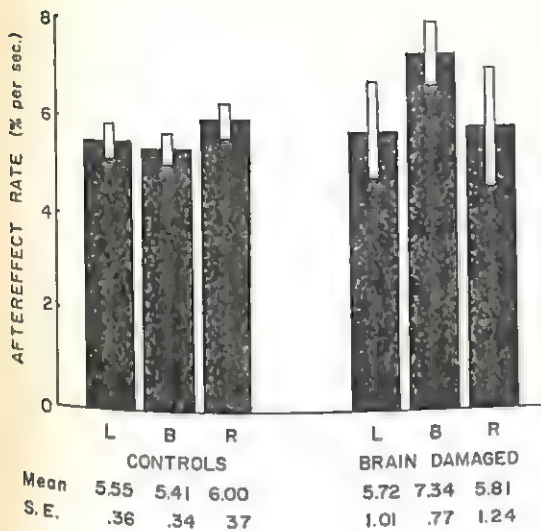


FIG. 2. Aftereffect rate in brain damaged and control subjects shown separately for the left-right interocular transfer condition (L), the binocular condition (B), and the right-left transfer condition (R).

than by the binocular condition; while in the brain damaged group more aftereffect was produced by the binocular condition. Table 1 makes possible a chi square test of this effect. In Table 1 a subject was classified in the left-hand column if both of his transfer aftereffects exceeded his binocular aftereffect. Subjects whose binocular aftereffect had a value intermediate between the two transfer values were classified in the second column. If both transfer aftereffects were less than the binocular aftereffect, the subject was assigned to the third column of the table. The overall chi square of 16.20 is significant beyond the .001 level of confidence. The evidence is therefore very strong that the brain damaged subjects responded to the transfer conditions in a manner different from that of the control subjects. It is remarkable that only two of the 51 brain damaged subjects had binocular aftereffect which was intermediate between the two transfer values. Brain damage appears to increase markedly the discrepancy between the binocular and the transfer conditions. In some cases the binocular aftereffect is greater, and in some cases the transfer aftereffect is greater. Inspection of the diagnoses in these two subgroups failed to reveal any systematic differences.

The data of Table 1 can also be analyzed for the separate groups. In the brain damaged group, for example, one would expect that 17 subjects would be grouped in each column under the null hypothesis that the transfer conditions produced no differential effects. Stated in slightly different form, this null hypothesis is that the two transfer measures and the binocular measure were drawn randomly from the same population. Under this assumption, the probabilities of each of the six permutations of  $L > R > B$  are equal. Each column of the table represents a pair of these six permutations and thus represents a probability of  $1/3$ . The chi square associated with this hypothesis is 20.59 which is significant beyond the .001 level of confidence. A similar analysis can be performed for the control subjects. The chi square for the control group is 3.29 and has a probability slightly less than .20, thus failing to provide strong statistical evidence for the differential effects of transfer in normals. The significance of the overall chi square is thus seen to be almost entirely attributable to the departure from chance expectancy of the brain damaged group. It is interesting to note that all previous investigators who have observed transfer effects have held that the magnitude of the transfer aftereffect is less than the binocular aftereffect. The present study provides the first quantitative information on this question and indicates that in normals transfer aftereffect is *not* smaller and is possibly larger than the binocular aftereffect.

TABLE 1

DISTRIBUTION OF BRAIN DAMAGED AND CONTROL PATIENTS ACCORDING TO RELATIVE MAGNITUDE OBTAINED UNDER TWO TRANSFER CONDITIONS AND A BINOCULAR CONDITION

Group	B < T	B ~ T	B > T	Total
Brain damaged	22	2	27	51
Control	17	12	8	37
Both groups	39	14	35	88

$$\chi^2 = 16.2035^*$$

Note.—B < T = Binocular less than transfers; B ~ T = Binocular intermediate between transfers; B > T = Binocular greater than transfers.  
\*  $p < .001$ .



TABLE 2  
MEAN VARIABILITY OF AFTEREFFECT

Group	L-R Transfer condition		Binocular condition		R-L Transfer condition	
	<i>M</i>	$\pm 1 SE$	<i>M</i>	$\pm 1 SE$	<i>M</i>	$\pm 1 SE$
Brain damaged ( <i>N</i> = 51)	23.62	$\pm 3.50$	17.69	$\pm 2.41$	17.59	$\pm 2.58$
Control ( <i>N</i> = 37)	3.23	$\pm .28$	3.13	$\pm .29$	4.14	$\pm .34$

In addition to the magnitude of the aftereffect, the technique of this study makes possible the measurement of variability of the aftereffect within a given series of trials. The measure of variability used was the variance of the 75 rates within a condition. Table 2 shows the average variance obtained from both groups under three different conditions. Inspection of Table 2 reveals at once that the brain damaged subjects displayed considerably more variability of aftereffect rate than did the controls. Again, there is heterogeneity which prevents the use of a parametric statistic. Statistical tests of the significance of differences in Table 2 will be approached in the discussion of diagnostic utility which follows.

On the basis of data from a smaller group of subjects, three criteria for diagnosing presence or absence of brain damage were set up. Criterion A classifies a subject as brain damaged if his aftereffect under any one or more of the three conditions exceeds a value of 10% per second. Criterion B classifies as brain damaged a subject who, at any time after the first trial, adjusts the stepping relay on the "wrong" side of the

zero position, i.e., in a way suggesting aftereffect in the same direction as the eliciting motion. Criterion C classifies a subject as brain damaged if his variance under any condition is greater than 10. Tables 3, 4, 5, and 6 contain contingency tables which show the number of subjects correctly and incorrectly classified by each criterion separately and by combining all three criteria. All chi squares are highly significant showing that departure from normal limits on any of the criteria is associated with presence or absence of brain damage.

TABLE 4  
DISTRIBUTION OF BRAIN DAMAGED AND CONTROL SUBJECTS ACCORDING TO CRITERION B: OCCURRENCE OF AFTEREFFECT IN THE SAME DIRECTION AS ELICITING MOTION

Group	Absent	Present	Total	Hit rate
Control	36	1	37	.97
Brain damaged	21	30	51	.59
Both groups	57	31	88	.75

$$\chi^2 = 23.17^{**}$$

$^{**} p < .0005.$

TABLE 3

DISTRIBUTION OF BRAIN DAMAGED AND CONTROL SUBJECTS ACCORDING TO CRITERION A: MAGNITUDE

Group	Less than 10% per second	More than 10% per second	Total	Hit rate
Control	36	1	37	.97
Brain damaged	25	26	51	.51
Both groups	61	27	88	.70

$$\chi^2 = 21.27^{**}$$

$^{**} p < .0005.$

TABLE 5  
DISTRIBUTION OF BRAIN DAMAGED AND CONTROL SUBJECTS ACCORDING TO CRITERION C: VARIANCE

Group	Less than 10% per second	More than 10% per second	Total	Hit rate
Control	37	0	37	1.00
Brain damaged	11	40	51	.74
Both groups	48	40	88	.88

$$\chi^2 = 50.09^{**}$$

$^{**} p < .0005.$

TABLE 6

DISTRIBUTION OF BRAIN DAMAGED AND CONTROL SUBJECTS ACCORDING TO ALL CRITERIA COMBINED

Group	Negative	Positive	Total	Hit rate
Control	35	2	37	.95
Brain damaged	3	48	51	.94
Both groups	38	50	88	.94

$$\chi^2 = 65.19^{**}$$

$^{**} p < .0005.$

## DISCUSSION

The results of the present study confirm both the findings based on persistence time and those based on the SAET. The persistence time studies (Spivack & Levine, 1957, 1959; Truss & Allen, 1959) predict larger aftereffects among brain damaged subjects. Table 3 shows that roughly half of the brain damaged subjects did have aftereffect which was greater than that experienced by virtually all of the controls.

The SAET studies suggest that the aftereffect is abnormally slight or missing in brain damaged subjects. Table 4 shows that approximately half of the brain damaged subjects had aftereffect which was either so slight, or of opposite direction, that they occasionally, or frequently, adjusted the stepping relay on the "wrong" side of zero, thus confirming the SAET results.

In comparing MMG results with SAET results it should be remembered that there are three major differences between the two techniques. First, the MMG measures the actual speed of the aftereffect and thus discovers those patients in whom the aftereffect is abnormally great. Second, the MMG measures over a much longer period and gives more opportunity for abnormality to be detected. Third, the spiral and rotation speed used in the MMG elicit almost twice the aftereffect elicited by the SAET (Scott, 1962). With respect to subjects having abnormally slight aftereffect, it may be that Criterion B fails in a number of cases precisely because the MMG does elicit more aftereffect than does the SAET. However, this possibility is offset to some extent by the much larger sample of the subject's aftereffect afforded by the MMG and the

greater variability of the brain damaged subject.

Thus far it has been assumed that the sometimes highly erratic performance of the brain damaged subject reflects irregular changes in his aftereffect rate. Another possible interpretation is that the brain damaged subject becomes confused, makes "errors," or perseverates. Such an interpretation could account for most of the findings reported above. There are clear reasons for doubting that this interpretation is correct. First, no subject was included in the study who failed to give conclusive evidence during the discrimination trials that he understood and could perform the task correctly. The possibility remains, of course, that the subject became confused or began to perseverate later, during the actual test. The subject's response was both verbal and motor. If the subject became confused, it might be expected that his push-button response and his verbal response would disagree. This very seldom happened, and a correction was always made whenever it did occur. In addition, there is the significant interaction of groups by conditions (Table 1). This interaction might be interpreted to mean that the introduction of the shutter for the transfer conditions simply increased the confusion of the brain damaged subjects so that their transfer results were deviant from their binocular results. If this were so, the variance measures obtained on the transfer conditions for the brain damaged subjects ought to be significantly inflated above those obtained under the binocular condition. Inspection of Table 2 does not suggest that any significant inflation of variance occurred.

The data seem to indicate, therefore, that aftereffect rate can be measured in brain damaged subjects and that the findings of the present study are based on perceptual impairment rather than general confusion or perseveration. The hit rates obtained with the MMG compare favorably with the highest hit rates reported for the SAET and are considerably higher than those obtained in the majority of SAET studies.

Eysenck, Holland, and Trouton (1957) have recently put forward a theory based on the concept of reactive inhibition. Accord-



ing to their theory, the brain damaged patient should have more reactive inhibition than the normal person. Reactive inhibition is supposed to coneract the aftereffect process. They predict, therefore, less aftereffect and shorter persistence of aftereffect for brain damaged subjects. The present findings fail to support this prediction if it is to be applied to brain damage in general, since the average aftereffect rate for brain damaged subjects was actually greater than for normals.

An important controversy in the area of motion aftereffects has always been the locus of the mediating structures. Wolgemuth (1911), for example, thought that the phenomenon was cortical, while Hunter (1914, 1915) advanced a purely retinal theory. Actual measurement of speed of transfer effects and the discovery that they are at least as great as the binocular effects casts much doubt on theories of retinal mediation. Also, the greater degree of disruption of transfer effects in brain damaged subjects than in controls seems more consistent with theories of central mediation than with retinal theories.

Methodologically speaking, the present study reiterates the superiority of objective, quantitative methods over more observational approaches. It illustrates the hazards inherent in the popular practice of relating gross behavioral observations to clinical variables. Often, as in the present case, findings based upon gross observations may be misleading because the finer detail is screened from the investigator's view.

In conclusion, a word of caution seems appropriate. Studies like the present one should perhaps be thought of as exploratory studies aimed at improving understanding of brain function rather than simply attempting to diagnose brain damage. In fact, since the brain mediates behavior of all kinds, it would not be surprising to find that almost any behavioral measure discriminates between brain damaged and normal individuals. The importance of studies of this kind lies rather in the fact that different behavioral measures are affected to different degrees and in different ways by damage to the central nervous system. Only through the systematic explora-

tion of such behavioral measures can significant contributions be made to the understanding of brain function. Clinical application should be accompanied by caution and the realization that much further work is needed in this area.

#### REFERENCES

- CANN, M. A. The negative after-effect of motion as a function of test stimulus texture. Unpublished doctoral dissertation, Boston University, 1961.
- DAVIDS, A., GOLDENBERG, L., & LAUFER, M. W. The relationship of the Archimedes spiral aftereffect and the trail making tests to brain damage in children. *J. consult. Psychol.*, 1957, 21, 429-433.
- DURUP, G. Le problème des impressions des mouvements consécutives d'ordre visuel. *Année psychol.*, 1928, 29, 1-56.
- EYSENCK, H. J., HOLLAND, H., & TROUTON, D. S. Drugs and personality: The effect of stimulant and depressant drugs on visual after-effects. *J. ment. Sci.*, 1957, 103, 650-655.
- FREEMAN, E., & JOSEY, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, 62, 794-796.
- FUCHS, F. Experimentelle Studien über das Bewegungsnachbild. *Z. Psychol.*, 1928, 106, 267-315.
- GALLESE, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254-258.
- GILBERSTADT, H., SCHEIN, J. D., & ROSEN, A. Further evaluation of the Archimedes spiral aftereffect. *J. consult. Psychol.*, 1958, 22, 243-248.
- GOLDBERG, L. A., & SMITH, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, 22, 153-157.
- HOLLAND, H. C., & BEECH, H. R. The spiral aftereffect as a test of brain damage. *J. ment. Sci.*, 1958, 104, 466-471.
- HUNTER, W. S. The aftereffect of visual motion. *Psychol. Rev.*, 1914, 21, 245-277.
- HUNTER, W. S. Retinal factors in visual after-movement. *Psychol. Rev.*, 1915, 22, 479-489.
- LONDON, P., & BRYAN, J. H. The influence of instructions on spiral aftereffect report. *Amer. Psychologist*, 1958, 13, 335. (Abstract)
- PAGE, H. A., RAKITA, G., KAPLAN, H. K., & SMITH, H. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, 21, 89-91.
- PRICE, A. C., & DEABLER, H. Diagnosis of organicity by means of the spiral aftereffect. *J. consult. Psychol.*, 1955, 20, 299-302.
- SCOTT, T. R. Summary of principal findings on a visual aftereffect of motion and brain pathology. *Newsltr. Res. Psychol.*, 1962, 4(3), 38-46.
- SCOTT, T. R., & MEDLIN, R. E. Measurement of rate of change of perceived size in the spiral

- aftereffect. *Amer. Psychologist*, 1960, 15, 420-421. (Abstract)
- SCOTT, T. R., & MEDLIN, R. E. Psychophysical measurement of the spiral aftereffect: The MMG. *Amer. J. Psychol.*, 1962, 75, 319-321.
- SINGER, J. G. The relationship between extent of aftereffect and speed of a rotated spiral. Unpublished doctoral dissertation, Boston University, 1959.
- SPIVACK, G., & LEVINE, M. The spiral aftereffect and reversible figures as measures of brain damage and memory. *J. Pers.*, 1957, 25, 767-778.
- SPIVACK, G., & LEVINE, M. Spiral aftereffect measures of satiation in brain-injured and normal subjects. *J. Pers.*, 1959, 27, 220-227.
- STANDLEE, L. S. The Archimedes negative aftereffect as an indication of memory impairment. *J. consult. Psychol.*, 1953, 17, 317.
- STERN, A. The latency of the spiral effect and aftereffect as a function of illumination and speed of rotation. Unpublished doctoral dissertation, University of Connecticut, 1959.
- STILSON, D. W., GYNTHIER, M. D., & GERTZ, B. Basal rate and the Archimedes spiral illusion. *J. consult. Psychol.*, 1957, 21, 435-437.
- SZILY, A. v. Bewegungsnachbild und Bewegungskontrast. *Z. Psychol. Physiol. Sinnesorg.*, 1905, 38, 81-154.
- SZILY, A. v. Zum Studium des Bewegungsnachbildes. *Z. Psychol. Physiol. Sinnesorg.*, 1907, 42, 109-114.
- TRUSS, C. V., & ALLEN, R. M. Duration of the spiral after-effect in cerebral palsy. *Percept. mot. Skills*, 1959, 9, 216-218.
- WOHLGEMUTH, A. On the aftereffect of seen movement. *Brit. J. Psychol., monogr. Suppl.*, 1911, 1, 1-117.

(Received May 17, 1962)



## PERCEPTUAL MODES AND ASTHMATIC SYMPTOMS: AN APPLICATION OF WITKIN'S HYPOTHESIS<sup>1</sup>

GERALD M. FISHBEIN<sup>2</sup>

*Veterans Administration Hospital, Brooklyn, New York*

The problem of tying personality formulations described by Witkin to behavioral correlates was investigated. The Rod and Frame Test and Embedded Figures Test were administered to 60 chronic asthmatic children who had been removed from their home environment. Rate of symptom alleviation following environment change and mode of perception were identified. Results indicated that rate of alleviation and length of time in a therapeutic milieu varied independently from perceptual mode. Also, asthmatic Ss tended to be more field dependent than the nonasthmatic standard. The negative results supported the contention that mode of perception may be an artifact of the experimental situation. The trend toward greater perceptual dependency of the asthmatics does lend some credence to both the description of the asthma personality and to Witkin's hypotheses.

Witkin, Lewis, Hertzman, Machover, Bretnall-Meissner, and Wapner (1954) published the results of a long-term study dealing with the relationship of personality traits to perception of the vertical. Depending on the accuracy of perception, the subjects were classified along a continuum called perceptual dependence-independence. The extreme groups had projective personality examinations administered to them (Witkin et al., 1954, p. 469).

Having described the projective test results characterizing each mode of perception, the problem of tying the abstract personality description to behavioral correlates remains. The present study was concerned with this problem, especially the passivity-activity dimension.

It was assumed, for the purpose of research, that perceptually dependent individuals "tend to be characterized by passivity in dealing with the environment." If the assumption is tenable, then a psychosomatic symptom exhibited by perceptually dependent subjects, which owes its genesis at least in part to environmental stimuli, ought to be partly alleviated when the environmental stimuli generating the symptom are removed. On the

other hand, independent or analytical perceptual performers who are more independent in relation to their environment, and who have the same and relatively equally severe psychosomatic symptom, ought to be less affected when a similar change from the pathogenic environment is made.

The symptom chosen for study was bronchial asthma. This disease has been recognized from the time of Hippocrates as being heavily influenced by emotional factors. Although in recent years the basic immunologic mechanisms in asthma have been described, current medical opinion acknowledges the importance of psychological factors. The subjects utilized in this study were chronic, severe, asthmatic children in whom psychological factors were thought to be primarily responsible for the chronicity of their asthma.

A further hypothesis considered was the contention that mode of perceiving is a basic character trait, and, therefore, probably resistant to change. In the subjects studied, strong pressures toward emotional maturity exist as a result of separation from the family, of group living, and psychotherapy. The effects of length of time in this combined medical and psychological residential treatment program were considered.

The final question concerned developmental changes described by Witkin. He has indicated that a progression from perceptual dependency to independency, reflected in

<sup>1</sup> Based on a dissertation submitted in partial fulfillment of the requirements for the Doctor of Philosophy Degree at the University of Denver, 1958.

<sup>2</sup> Formerly, Jewish National Home for Asthmatic Children, Denver, Colorado.

the experimental tasks, is a part of normal development, characterized by a predictable pattern. There is reason to believe, however, that asthmatic children might represent a selected range of the distribution of modes of perception. The particular symbiotic relationship between the asthmatic child and his mother, often described in the literature (Fine, 1948; French & Alexander, 1941; Miller & Baruch, 1951) should predispose to perceptual dependency. The problem for this research was to determine whether the developmental changes described by Witkin for normal children obtained for the present sample, and whether the range of perceptual mode differs in the asthmatic as compared with normal subjects.

#### METHOD

The Rod and Frame and the Embedded Figures Test were administered to 60 children, ages 9 to 15, who were patients at the Jewish National Home for Asthmatic Children (JNHAC) Denver, Colorado. They had been removed from their home environment in an attempt to alleviate their chronic asthma. The subjects in residence from 9 to 18 months were selected.

In the total group from which subjects were selected, 50% were free of asthma upon or within a few days after admission. In another 30% there was progressive lessening of asthma during the first year and freedom from asthma during the second and last year of residence. An additional 10% lost their asthma upon admission but the condition recurred during the pollen season. In still another 5% of the patients asthma continued unchanged but did not interfere with most usual activities.

The remaining 5% continued to have asthma unaffected by institutionalization. The factors thought to be primarily responsible for the patients' improvement were the separation from their parents, the group living experience which appears to foster independence, and brief individual psychotherapy.

The subjects were assigned to one of three groups depending on the severity of their conditions. Group I included those patients who had rapid change of symptoms following arrival at the Home. There was no asthma present except for infrequent attacks which were easily controllable. Group II included those patients whose improvement following arrival was slower than Group I. There was continued asthma intensified by hay fever, pollen asthma, upper respiratory infection, or a combination of these. They occasionally required hospitalization. Group III comprised the most severe group; almost all of the children in this group were taking continued steroid (Sterane: Pfizer & Co., New York) medication. Hospitalization was often required.

Selection of the groups was made by utilizing as an index of severity cutoff points on a distribution of total required nebulizations, an aerosol therapy used to abort an asthmatic attack. The cutoff points adhered to the natural breaks in the distribution as closely as possible. In addition, opinions of the attending physician, based on severity, duration, and seasonal pattern of the attacks, and frequency of hospitalization were considered. The three groups were matched for age, sex, intelligence, and length of time, in months at the JNHAC. An analysis of variance computed for each of these variables failed to reveal significant difference. Table 1 presents these matching data.

The Rod and Frame apparatus and the Embedded Figures Test used in the research were supplied by Witkin's laboratories. With the exception of two changes, the procedure followed was that described by Witkin. The changes made were the elimination of the Rod and Frame body tilted trials, and the shortening of the Embedded Figures Test to 12 figures in order to maintain the subject's interest and motivation. Since each part of the Rod and Frame Test correlates highly with the total, the body erect position alone adequately reflects the subject's mode of perception. A study by Jackson (1956) reported that the 12-figure form of the Embedded Figures Test correlated .99 with the full test in Jackson's groups. It was, therefore, felt that shortening of the tests was justifiable.

A pilot study was run to determine whether those children taking steroid therapy perceived the Rod

TABLE 1  
VARIABLES USED IN MATCHING SEVERITY GROUPS

Group	N	Sex		Age		IQ <sup>a</sup>		Length in months in JNHAC	
		M	F	M	SD	M	SD	M	SD
I	27	18	9	12.60	2.1	101.94	16.51	13.11	4.31
II	23	15	8	12.39	1.97	101.72	19.12	13.69	3.66
III	10	6	4	13.00	2.05	107.55	21.18	15.00	5.25

<sup>a</sup> Wechsler Intelligence Scale for Children or Wechsler Bellevue Form I.



and Frame differently from those not on steroids. All subjects who had been previously selected and who had not taken steroids during their residence were pooled. From this group 20 subjects were selected using a table of random numbers. Ten subjects were then assigned at random to either an experimental or control group by another staff psychologist of the Home. Tabulation of the data revealed that the experimental and control groups did not differ in age, sex, intelligence, or length of institutionalization.

The experimental group received 10 milligrams of the steroid medication, orally administered by the charge nurse, in the same manner as other drugs in the Home. The dosage represented the mean amount received by members of Group III. The control group received a placebo. Retesting after a mean time of 8.1 and 8.2 days, respectively, for experimental and control groups was done only with the Rod and Frame. It was felt that practice effect after a short time on the Embedded Figures Test would make the results difficult to evaluate.

### RESULTS

The results of the pilot study failed to show a significant effect of one administration of 10 milligrams of steroid on Rod and Frame performance. The  $t$  values between means for experimental and control groups before steroid or placebo, after steroid or placebo, and between the means of experimental and control groups were not statistically significant.

Since it was also desired to determine whether continued steroid dosage has an effect on Rod and Frame performance, a group of subjects from Group III was selected and matched on the relevant variables with the experimental group of the pilot study. It had already been observed that the administration of 10 milligrams of steroid did not materially affect Rod and Frame performance. If Rod and Frame presteroid performance did not differ significantly from the performance

of a matched group who regularly were treated with steroid, then the probability would be very high that steroid medication continued for 1 year also has no significant effect on the Rod and Frame score. The  $t$  of .58 for the comparison failed to reach significance ( $.60 < p < .70$ ,  $df = 16$ ).

*Hypothesis 1.* The results essentially fail to support the hypothesis that following removal from the home environment, the asthmatic symptoms of perceptually dependent children will be alleviated more rapidly than perceptually independent children.

Each raw score in the Rod and Frame and Embedded Figures Test distribution was converted to a  $Z$  score following Witkin's procedures. The two  $Z$  scores for each subject were averaged, and a rank in the total mode of perception distribution assigned ( $N = 58$ ). Since the distribution of these scores were skewed, the assumption of normality of the population was suspect. Therefore, a Kruskal-Wallis analysis of variance following the procedure outlined in Siegel (1956) was computed. The  $H$  value of .029 failed to indicate significant differences ( $.80 < p < .90$ ) in the obtained ranks on the mode of perception distribution for those subjects who were diagnosed as being different in their symptomatology. However, the mean scores for each severity group on the Embedded Figures Test were in the predicted direction. On the Rod and Frame, two of the three means were also in the predicted direction. These data are presented in tabular form in Table 3. For the Rod and Frame, the mean represents the mean error in degrees in setting the rod to the vertical. The mean for the Embedded Figures Test is the mean number of seconds required by the subjects to locate the simple figure in the complex one.

Since the data was generally in the predicted direction and since the variability was greater than expected using Witkin's subjects as a frame of reference, an attempt to lower the variability was made. This was done by analyzing means based only on the last four trials of the Rod and Frame, so that the first four trials could logically be considered practice trials for each subject. This also failed to result in significance. Similarly, an attempt to dichotomize, instead of the initial attempt

TABLE 2

EFFECT OF 10 MILLIGRAMS STERANE ADMINISTRATION ON ROD AND FRAME SCORES

Group	N	Pre-Sterane		Post-Sterane	
		$M^a$	$SD$	$M$	$SD$
Experimental	9	15.93	13.12	11.73	10.08
Control	10	14.85	14.28	7.79	7.14

<sup>a</sup> Means are mean error in degrees in setting the rod to the vertical.

TABLE 3  
COMPARISON OF SEVERITY GROUPS ON THE ROD AND FRAME AND  
EMBEDDED FIGURES TESTS

Group	Change in Symptoms					
	I: Rapid		II: Intermediate		III: Little	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Rod and Frame	12.88	10.70	11.39	10.45	15.14	10.00
Embedded-Figures	143.27	80.68	136.80	85.90	128.64	82.80
<i>N</i>	27		23		10	

to trichotomize, the severity groups also failed to result in significant differences between the groups.

*Hypothesis II.* This prediction stated that a relationship exists between length of residential treatment (milieu therapy) and modes of perception. The subjects, all of whom already had been tested, were assigned to one of two groups, depending on their length of residence. The mean number of months in residence for Group A, the more recent arrivals, was 9.33 with a standard deviation of 1.97. For the second group, Group B, the mean number of months in residence was 16.21, with a standard deviation of 1.93. The two groups were matched for severity of illness, sex, and intelligence. They differed significantly, however, in age. The children of Group B were almost 1 year older as a group than those of Group A.

The means and standard deviations obtained by Group A were as follows: Rod and Frame,  $M = 13.9$ ;  $SD = 10.83$ ; Embedded Figures Test,  $M = 141.64$ ;  $SD = 92.02$ . The means and standard deviations obtained by Group B were: Rod and Frame,  $M = 11.79$ ;  $SD = 10.30$ ; Embedded Figures Test,  $M = 136.99$ ;  $SD = 75.23$ . In the evaluation of these data, an analysis of covariance was computed which statistically controlled the age factor following the procedure described by McNemar (1955). The resulting  $F$  yielded a probability greater than .05. It cannot be concluded, therefore, that there are significant differences between the perceptual scores of the groups as a function of length of residence in the Home.

*Hypothesis III.* This prediction was that the developmental change described by Witkin

obtained for the present sample. Also, that these asthmatic children would be significantly more field dependent than the normals.

To answer these questions the subjects were divided into four age groups, similar in intelligence. Comparisons of means and standard deviations for each age group on the Rod and Frame and Embedded Figures Test were made. This analysis suggested a progressive decrease in the extent of field dependence with increasing age. The obtained data were then compared with data reported by Witkin et al. (1954, pp. 122, 124). The similarity in the shape of the two distributions for both the Rod and Frame and the Embedded Figures Test was striking. The distributions of the obtained data, compared with the normative data, only tended to differ significantly ( $.04 < p < .10$ ) in the extent of field dependence. The asthmatic subjects were more field dependent.

#### DISCUSSION

While it appears that asthmatic children tend to be more perceptually dependent than nonasthmatic children, the fluctuation of their symptoms appears to be independent of their mode of perception. The hypotheses were based on Witkin's description of modes of perceiving as a "coping mechanism." That is, it represents an individual's characteristic way of coping with the stress of his environment as well as the way in which he manages his "inner life." In a somewhat similar manner, the case has been made for asthma as a coping mechanism. The psychogenic aspects of the disease have been described as the asthma-prone individual's reaction to various types of environmental stress ranging from a



nonspecific variety to the stress resulting from fear of separation from a rejecting mother. If the coping mechanism reasoning is accurate, the hypothesized relationships should have been revealed.

A possible explanation for the negative results is suggested by Gibson (1952). He asserts that the individual differences found by Witkin are probably a function of the experimental situation, rather than demonstrating that individuals differ in the basic ways they perceive space.

Gibson's explanation could also be applied to the failure to achieve significant differences concerning Hypothesis II. Although it cannot be ruled out that no basic personality change has occurred during residence, there is very strong evidence that the children do increasingly manifest more traits of greater psychological maturity with length of residence. They assume responsibilities far removed from any they had assumed in their home environment. They often change from timid, whining individuals to independent, assertive children. Symptoms such as manifest anxiety, enuresis, and infantile speech frequently improve noticeably. However, the evidence that asthmatic children are more perceptually dependent than nonasthmatics is in agreement with

the general personality structure drawn for the asthmatic. The trend toward greater perceptual dependence in the asthmatics would lend credence to Witkin's hypotheses.

#### REFERENCES

- FINE, R. A quantitative study of personality factors related to bronchial asthma in childhood. Unpublished doctoral dissertation, University of Southern California, 1948.
- FRENCH, T., & ALEXANDER, F. Psychogenic factors in bronchial asthma. *Psychosom. Med. Monogr.*, 1941, 1(4, Pt. 1).
- GIBSON, J. J. The relation between visual and postural determinants of the phenomenal vertical. *Psychol. Rev.*, 1952, 59, 370-375.
- JACKSON, D. N. A short form of Witkin's Embedded Figures Test. *J. abnorm. soc. Psychol.*, 1956, 53, 254-255.
- MCNEMAR, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.
- MILLER, H., & BARUCH, DOROTHY, W. Maternal rejection aspects in the treatment of bronchial asthma. In H. A. Abramson (Ed.), *Somatic and psychiatric treatment of asthma*. Baltimore: Williams & Wilkins, 1951. Pp. 632-654.
- SEIGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., BRETNALL-MEISSNER, P., & WAFNER, S. *Personality through perception*. New York: Harper, 1954.

(Received May 31, 1962)

## MANIFEST ANXIETY, SOCIAL DESIRABILITY, OR RESPONSE SET

HENRY E. ADAMS AND ALBERT C. KIRBY

*Louisiana State University*

The present investigation tested Edwards' hypothesis that the *SD* scale is not confounded with response set and determined the extent that response set contaminates Taylor's *MA* scale. By reversing the items of these scales, true and false scales for each personality scale were obtained. Statistical procedures were then utilized to determine response set and content scores for each scale. The results indicate that the *MA* scale is relatively free of response set and that the *SD* scale is measuring both response set and social desirability. It should not be assumed automatically that a scale which has items stated in a single direction is measuring response set or that a scale with items stated in both directions is not.

Responses to items of personality inventories are frequently determined by the individual's desire to place himself in a favorable light. Edwards (1953, 1957) calls this tendency the social desirability-undesirability dimension and has constructed a scale (the Social Desirability scale) (*SD* scale) which measures this tendency of the individual to give socially desirable response in personality inventories. He has presented evidence that the *SD* scale correlated quite highly with other personality scales, either positively or negatively depending on whether the trait measured is socially desirable or undesirable. As a result Edwards has stated that all personality scales are contaminated with the social desirability variable.

It has been pointed out by Cronbach (1946, 1950) that personality inventories are also confounded with response sets. Two of these response sets are the tendency to acquiesce (respond "true") or to dissent (respond "false"). On the basis of this hypothesis it is possible that the *SD* scale is actually measuring response set instead of the social desirability variable since 30 of the 39 items in the *SD* scale are keyed "false." Edwards (1957) has discussed this possibility and has tested between these two hypotheses.

However, Edwards seems to have made the assumption that all the personality scales were equally contaminated with response set. This is probably not the case. Thus, he failed to differentiate between content scores and response set scores of each scale. Webster

(1958) and Helmstadler (1957) have presented a statistical procedure for obtaining separate set and content components of a scale. The present investigation is an attempt to further test between the response set and social desirability hypotheses by using these statistical procedures and by reversing the statement of the items of the *SD* scale in order to obtain a true and false scale for social desirability. This procedure is described by Bass (1955) in his investigation of the *F* scale. Further, this procedure will also be conducted with the Manifest Anxiety scale (*MA* scale) (Taylor, 1953) in order to determine if the relationship between the *SD* scale and the *MA* scale is due to response set, social desirability, or an interaction of these two variables.

### METHOD

*Scales.* The 39-item *SD* scale and the 50-item *MA* scale were utilized. The statement of the items in both scales were reversed so that items stated in the positive direction were stated in the negative direction and vice versa. A negative and a positive scale for Manifest Anxiety and Social Desirability were then constructed, making a total of four scales. These scales are not identical with the original *MA* or *SD* scale in that the original scales of both contained positive and negative items. The reversals were not particularly difficult to construct since the items in these scales are fairly concise and directly stated. This appeared to be particularly true of the *MA* scale. The criterion for adequate reversal of the meaning of an item was agreement of three judges.

*Subjects and procedures.* The subjects were 22 male and 29 female college students at Louisiana State University. They were given an *SD* scale (either positive or negative) and an *MA* scale (either



positive or negative) in the initial session and reversed scales 2 weeks later. A counterbalanced design was used to control effects of order of presentation of *MA* and *SD* scales and the effects of order of presentation of positive and negative scales. The male and female subjects were pooled since sex differences were not relevant to this study. Further, Edwards (1957) and Taylor (1953) both found no significant differences on these scales between sexes.

*Data analysis.* Correlations between all scales were computed from the raw scores. Response set (*S*) and content scores (*C*) were also computed for *SD* and *MA* using essentially the formula described by Messick and Frederiksen (1958, p. 689) which is derived from the simplified ratio formula of Helmstadler (1957). The equations are as follows:

$$C = F_{T/N_T} + U_{F/N_F} \quad [1]$$
$$S = F_{T/N_T} - U_{F/N_F} = (2 - C) \quad [2]$$

*N<sub>T</sub>* and *N<sub>F</sub>* are numbers of items keyed true and false respectively. *F<sub>T</sub>* is number of true items agreed with by the subject. *U<sub>F</sub>* is number of false items disagreed with. The content score varies from 0 to 2 while the set scores vary from +1 to -1.

RESULTS AND DISCUSSION

Table 1 shows the means and standard deviations for various scales. The response set mean for the *MA* scale is close to zero, indicating that response set in this scale appears to be mainly random error. This does not seem to be true for the *SD* scale where the mean of the response set scores is much higher. It is also apparent that reversing the original scales tends to systematically increase the means and standard deviations of raw scores.

As may be seen in Table 2, the correlation between the *MA* true and false scales is .62, which is significant at the .01 level; however, the correlation between the *SD* true and

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR RAW CONTENT, AND RESPONSE SET SCORES OF *SD* AND *MA* SCALES

Scales	<i>N</i>	<i>M</i>	<i>SD</i>
<i>MA</i> true	51	14.76	7.00
<i>MA</i> false	51	19.22	8.14
<i>SD</i> true	51	22.90	4.52
<i>SD</i> false	51	30.08	5.57
<i>MA</i> content	51	.68	.27
<i>MA</i> set	51	-.08	.13
<i>SD</i> content	51	1.33	.29
<i>SD</i> set	51	-.37	.22

TABLE 2  
CORRELATIONS OF RAW SCORES OF *SD* AND *MA* SCALES

	<i>MA</i> false	<i>SD</i> true	<i>SD</i> false
<i>MA</i> true	.62**	-.63**	-.73**
<i>MA</i> false	—	-.52**	-.80**
<i>SD</i> true	—	—	.32*

\* *p* < .05.  
\*\* *p* < .01.

false scale is only .32, which is significant at the .05 level. The correlation for the *SD* true and false scales is significantly smaller than the correlation obtained by Edwards and indicates that the *SD* scale is confounded with response set. These results indicate that the *SD* scale may not be measuring a unidimensional trait.

The correlations between the various *SD* and *MA* scales are significant at the .01 level and are in the direction predicted by the social desirability hypothesis. However, there are 22 overlapping items in these scales which spuriously inflates the correlations. Further, these correlations are not contradictory to the response set hypothesis in that the *MA* scales appear to be content scales, whereas the *SD* scales appear to be measuring both set and content. Thus the response set hypothesis could not predict the direction of the correlation of *SD* set scores with the *MA* scales since response set in the *MA* scales appears to be largely random error.

As may be seen in Table 3, the *MA* set scores are not significantly related to any of the other variables. These results would tend to support the conclusion, that response set in the *MA* scale is largely random error. However, the *SD* set scores are significantly re-

TABLE 3  
CORRELATIONS OF CONTENT AND SET SCORES OF THE *SD* AND *MA* SCALES

	<i>MA</i> set	<i>SD</i> content	<i>SD</i> set
<i>MA</i> content	-.02	-.86**	.66**
<i>MA</i> set	—	.00	-.05
<i>SD</i> content	—	—	-.68**

\*\* *p* < .01.

lated to the *SD* content scores and the *MA* content scores. This data would tend to support the conclusion that response set in the *SD* scales is not random error but systematic variance. The high negative correlation between the *MA* and *SD* content is as expected, due to the overlapping items.

If the items which do not overlap in the two scales are examined it seems fairly apparent why the *SD* scale is contaminated with response set while the *MA* scale is not. The items in the *MA* scale tend to be concise and direct while the items in the *SD* are somewhat more ambiguous. Ambiguity of items appears to be the major variable in determining the extent of response set in a personality scale. As the ambiguity of the items of a scale increases, the probability that a scale is confounded with response set increases. Another important variable is the individual ability to discriminate or tolerate ambiguity (Adams, 1961). Response set is likely to be an important factor in a scale when it is given to individuals whose ability to discriminate or structure the items is minimal.

In conclusion, Edwards' hypothesis that the *SD* scale is not confounded with response set is not supported by the present evidence. Further, the response set hypothesis does not necessarily predict positive correlations for scales scored in the same direction and negative correlations for scales scored in different directions. This prediction is true only when both scales are measuring response set. Prob-

ably the best procedure for determining whether a given scale is confounded with response set is the statistical techniques described by Helmstadler (1957) and Webster (1958) utilizing original and reversed items. This technique is a direct measure of response set in personality scales.

#### REFERENCES

- ADAMS, H. E. Authoritarianism, statistical rigidity, and a sampling control of response set. *J. Psychol.*, 1961, 53, 47-53.
- BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, 51, 616-623.
- CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.
- CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.
- EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, 37, 90-93.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- HELMSTADLER, G. C. Procedures for obtaining separate set and content components of a test score. *Psychometrika*, 1957, 22, 381-394.
- MESSICK, S., & FREDERIKSEN, N. Ability, acquiescence, "authoritarianism." *Psychol. Rep.*, 1958, 4, 687-697.
- TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WEBSTER, H. Correcting personality scales for response sets or suppression effects. *Psychol. Bull.*, 1958, 55, 62-64.

(Received June 6, 1962)



## A FURTHER APPRAISAL OF THE BODY BOUNDARY CONCEPT<sup>1</sup>

SEYMOUR FISHER

*Uplate Medical Center, State University of New York*

A series of studies were reviewed which cross-validated and extended previous findings concerning the relationship of two measures of body image boundary definiteness (barrier and penetration scores) to various levels of behavior. Support was found for the view that the more definite an individual's boundaries the more likely he is to manifest relatively higher physiological reactivity in body exterior as contrasted to body interior sectors. There was substantial evidence too that with increasing boundary definiteness there is greater ability to adjust adequately to disablement of one's body, to maintain normal ego integration, and to be effectively communicative in small group settings. Finally, the rationale was discussed for interpreting the barrier and penetration scores within a body image framework.

It is known that one assigns qualities (e.g., size, attractiveness) to one's body in terms of personalized standards which bear little relationship to actual body characteristics (Cleveland, Fisher, Reitman, & Rothaus, 1962; Fisher & Cleveland, 1958b; Lhermitte, 1935; Secord, 1953; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). When an individual perceives his own body he seems to become uniquely ego involved (Beloff & Beloff, 1957; Wolff, 1943) and he introduces systematic biases which may reveal a good deal about him. The term "body image" has been adopted to designate the attitudinal framework which defines the individual's long-term concept of his body and also influences his perception of it. Investigators have explored many body image dimensions with the intent of predicting from them to other behavioral variables.

Fisher and Cleveland (1958b) proposed that a fundamental aspect of the body image has to do with the manner in which an individual perceives his body boundaries. It was suggested that there is variation in how definite or firm one perceives one's body boundaries to be. Thus, the individual may view his body as clearly and sharply bounded, with a high degree of differentiation from non-self objects. But contrastingly, he may regard his body as lacking demarcation from what is "out there." The view was taken that the process of

learning to separate one's body from its environs is fundamental in the establishment of identity and that therefore the character of the body image boundary should provide important information about adjustment strategies. In translating the concept of body image boundaries into operational terms, Fisher and Cleveland developed a method which involves scoring the properties of the boundary regions of percepts elicited by ink blot stimuli. Boundary definiteness was found to be equivalent to the degree to which definite structure, substance, and surface qualities were assigned to the periphery of ink blot images. Responses such as the following were considered to represent an expression of definite boundaries: cave with rocky walls, man in armor, animal with striped skin, turtle with shell, mummy wrapped up, woman in fancy costume. These percepts in which the boundary is positively highlighted in some way are labeled "barrier responses." A second boundary index was also formulated which concerns ink blot percepts emphasizing the weakness, lack of substance, and penetrability of persons and objects. The term "penetration response" is applied to them, and some examples follow: mashed bug, person bleeding, broken body, torn coat, body seen through a fluoroscope. The barrier and penetration scores are both significantly correlated with the total number of responses given by the subject to a set of ink blot stimuli (e.g., Rorschach or Holtzman blots). Therefore, it is necessary to request

<sup>1</sup> This review was partially supported by United States Public Health Service Grant M-5761.

that the subjects produce a uniform number of responses for each blot. The interscorer reliability for evaluating barrier and penetration percepts varies from .82 to .97, with most values clustering in the high .80s and low .90s (Fisher & Cleveland, 1958b; Holtzman, Thorpe, Swartz, & Herron, 1961, Mausner, 1961;<sup>2</sup> Ramer, 1961).

In a book entitled *Body Image and Personality* Fisher and Cleveland demonstrated that the body image boundary scores could predict a variety of behaviors. Rather surprisingly, these scores proved to be significantly linked with such a range of phenomena as patterning of body sensations; differences in body exterior versus interior physiological reactivity; psychopathology; and conduct in small group situations. Since 1958 when *Body Image and Personality* was first published many studies have been undertaken by the original authors and also others to cross-validate and extend the findings that were obtained. The present paper seeks to summarize and integrate these studies. A series of circumscribed topical areas will be presented. Past findings pertinent to each will be described and followed by accounts of more recent work.

#### PSYCHOPHYSIOLOGICAL PATTERNS

Particular physiological patterns were among the first correlates of boundary definiteness to be observed. It had been initially noted that patients with rheumatoid arthritis, neurodermatitis, and conversion symptoms involving the musculature were characterized by higher barrier and lower penetration scores than patients with stomach ulcers or spastic colitis. From such findings the notion evolved that persons with definite boundaries who develop psychosomatic symptoms under stress tend to do so in the exterior body layers (viz., skin and muscle); whereas persons with indefinite boundaries manifest such symptoms in the interior body regions (viz., stomach, gut, and other internal organs). This exterior-interior model was later extended to persons in the normal range by findings which indi-

cated that normal subjects with definite boundaries manifest relatively high reactivity in the muscles and skin (e.g., in terms of GSR and EMG) and low reactivity at interior sites (exemplified by heart rate); but with just the obverse pattern appearing for those with indefinite boundaries.

Since the first presentation of these formulations a variety of studies has been undertaken to evaluate them further. Cleveland and Fisher (1960) and Fisher and Cleveland (1960) reported a replication of the original differences they observed in barrier and penetration scores between patients with rheumatoid arthritis and patients with stomach ulcers. Arthritics ( $N = 26$ ) exceeded patients with stomach ulcers ( $N = 34$ ) in barrier responses (.001 level) and in turn were exceeded by them in penetration responses (.10-.05 level). It was also found that the arthritics were characterized by a significantly lower heart rate (interior reactivity) and higher number of GSR responses (exterior reactivity) than the ulcer patients under stress conditions (both at .001 level).

Williams (1962) evaluated arthritic ( $N = 20$ ) and ulcer patients ( $N = 20$ ) and reaffirmed that the former had higher barrier (.10-.05 level) and lower penetration (.001 level) scores than the latter. In addition, he found significant trends for heart rate to be higher, and muscle potential lower, in ulcer as compared to arthritic patients under certain conditions. However, predicted differences in GSR were not observed.

Fitzgerald (1961) investigated children with Legg-Calvé-Perthes (LCP) disease. This "disease," whose cause is unknown, results in damage to the hip joint which is perhaps analogous to some phenomena encountered in rheumatoid arthritis. Fitzgerald examined LCP from the perspective of the body image boundary theory. He compared 20 children with LCP to 15 controls. As hypothesized, the LCP children proved to have significantly higher boundary definiteness than the controls. Also, the LCP subjects as compared to the control subjects were significantly more motorically expressive and displayed relatively greater skill in performance than verbal tests (WISC).

Eigenbrode and Shipman (1960) are the

<sup>2</sup> B. Mausner, unpublished progress report entitled "Experimental Studies of Social Interaction," 1961, National Institute of Mental Health, Grant No. M-2836.



only investigators to date who have failed completely to replicate the boundary distinction between psychosomatic patients with interior as opposed to exterior symptoms. They extracted from their clinical files the Rorschach protocols of 54 patients with "psychosomatic skin disorders" (exterior) and 29 patients with internal disorders (e.g., stomach ulcer, genito-urinary disease). Their scorings for the barrier and penetration variables revealed only chance differences between the two groups.

Malev (1961) extended the exploration of the exterior-interior hypothesis to normal children. His design involved 30 male 6-year-old and 30 male 8-year-old subjects. The subjects' mothers were interviewed to ascertain the frequency with which each subject had been characterized by exterior versus interior symptoms. In addition, GSR, heart rate, and blood pressure were recorded from the subjects under conditions of rest and stress. The data indicated that at both ages 6 and 8 the greater the boundary definiteness of the subjects the more likely they were to manifest a significant predominance of exterior over interior symptoms. Analysis of the physiological data demonstrated that in the 6-year-old group heart rate was significantly negatively related to boundary definiteness, but GSR bore only a chance relationship to it. In the 8-year-old group, all the physiological measures proved to be positively correlated with boundary definiteness. This pattern differs from that characterizing adults; and it led Malev to question whether the physiological correlates of boundary definiteness may not systematically differ at certain age levels from those found in adults. However, he did point out further that the symptom patterns reported by mothers in both 6- and 8-year-olds are correlated with the barrier and penetration scores in the same directions as they have been found to be in adults. He speculated on this basis that symptom reports by mothers about their children may be better indicators of long-term autonomic patterning than brief samples of physiological reactivity obtained in artificial laboratory situations.

Brown (1959) applied the barrier and penetration indices to the discrimination of

20 college students with contact dermatitis from 20 controls. He found a borderline tendency (.20-.10 level) for the dermatitis cases to have higher barrier and lower penetration scores than the controls. He noted that the experimental and control groups were "impure" insofar as the former probably contained people "exposed to powerful allergies" and the latter "young adults who may yet develop contact dermatitis."

Cleveland and Johnson (in press) compared the boundary attributes of 25 young men with coronary disease with a matched group of 25 men awaiting surgery. They discerned no differences in barrier scores between the groups. However, penetration responses were significantly higher in the coronary than in the presurgical patients.

Davis (1960) investigated reactivity in 25 men with unusually high barrier scores and 25 men with unusually low scores. He obtained electromyograph, skin resistance, blood pressure, and ballistocardiograph measures under rest and stress conditions. There were four differences during rest which were in the predicted direction but only two (EMG, stroke volume) were statistically significant. An analysis of changes in reactivity from rest to stress indicated that, as predicted, EMG was significantly greater in the high than the low barrier group and conversely that heart rate, stroke volume, and total cardiac output were of larger magnitude among low than high barrier subjects. A borderline tendency (.06 level) appeared for total peripheral resistance to be greater in high than low barrier subjects. Although no predictions had been made about blood pressure, it was found that during both rest and stress mean systolic blood pressure was more elevated in the low than high barrier groups. Davis considered his results to be supportive of the hypothesis that high and low barrier persons react differentially at exterior and interior body sites.

Fisher (1959b) sought to replicate Davis' findings in a population which was quite different in sex, age, and barrier score selectivity. Thirty girls with a median age of 14 were studied whose barrier scores were distributed over the entire range rather than representing extremes. GSR was taken as

an exterior response index and heart rate as an interior index. Recordings were secured during both rest and stress. The barrier score proved to be significantly positively correlated with GSR frequency and negatively correlated with heart rate during stress. Under rest conditions only a borderline positive correlation (.10-.05 level) between the barrier score and GSR frequency appeared.

It is appropriate at this point to refer to Lacey's formulation (in Rubinstein & Parloff, 1959)

that skin conductance increase is excitatory, whereas increase of cardiac rate is inhibitory of . . . transaction of the organism with the environment. The pattern of response obtained when recording skin resistance and heart rate may reveal occasions when the individual is "open" to his environment and ready to react to it, or conversely, when the individual is not "open" and indeed, instrumentally 'rejects' the environment (p. 205).

Lacey suggested on the basis of a review of experiments by Darrow (1929), Davis (1957) and Lacey (in Rubinstein & Parloff, 1959) that an attitude of being open and receptive to the world is accompanied by increased skin conductance and deceleration in heart rate, with the converse pattern typical of a closed unreceptive orientation. This formulation is obviously analogous to the inside-outside model which proposes contrasting levels of skin conductance and heart rate as boundary definiteness varies. Indeed, the analogy becomes even more precise if one considers that the definite boundary person (characterized by tendencies to high skin conductance and low heart rate) has been found to be more "open" and less defensive in dealing with the environment than the indefinite boundary person (Fisher & Cleveland, 1958b).

The tenor of the material presented in this section has been supportive of the boundary concept of exterior versus interior reactivity. Only one (Eigenbrode & Shipman, 1960) of numerous studies contradicted this model. Heart rate and muscle potential have fared best as measures, respectively, of interior and exterior reactivity. GSR and skin resistance have been inconsistent as indices of exterior reactivity and need to be evaluated further.

Also, in two exploratory studies blood pressure has proved to be significantly but variably linked with boundary definiteness (viz., positively in children and negatively in adults).

#### RESPONSE TO STRESS

The possession of definite boundaries was shown in *Body Image and Personality* (Fisher & Cleveland, 1958b) to permit the individual to deal relatively efficiently with stress. There was particular evidence that reaction to the stress of body disablement was likely to be less severe in the definite than indefinite bounded individual. For example, adequacy of adjustment to polio disablement and also amputation were positively correlated with the barrier score. Landau (1960) has provided support for these findings. She studied 40 paraplegic men with spinal cord injuries. Their adjustment to their disablement was evaluated by observational ratings and by means of a sentence completion test. It was found that the higher a patient's barrier score the better was his adjustment as defined by behavioral ratings (.01 level) and sentence completion responses (.01 level). The penetration score had only a chance relationship to the criteria. Landau noted that the barrier score was not related to the duration of time the patient had been disabled. This is congruent with earlier studies (Fisher & Cleveland, 1958b; Ware, Fisher, & Cleveland, 1957) in which the barrier score seemed not to be influenced either by the amount or duration of damage sustained by the body. Indeed, more recently Fisher (1959a) found that the decline in physique accompanying advanced aging does not result in a decrease in barrier responses.

McConnell and Daston (1961) considered the responses of 28 women to the stress occasioned by their own pregnancies. Subjects were seen pre- and postdelivery. Initially, each was given the Rorschach; the Osgood Semantic Differential to be applied to her own body; and a structured interview. Postdelivery, only the Rorschach and the Osgood Semantic Differential were repeated. The favorableness with which the subjects viewed their pregnancies turned out to be positively



manipulated by the experimenter so as to contradict the subject's reports. Each subject's responses to the contradictions were evaluated. It was found that the higher the barrier score the greater the likelihood that the subject would consider his own judgments, rather than his partner's, to be correct. Also, the barrier score was positively linked with the degree that the subject felt satisfied with his own performance and negatively related to his level of anxiety. Incidentally, the barrier score was significantly and negatively related to the *L* (Lie) score on the MMPI. This suggested that the individual with definite boundaries has less need to present an exaggerated favorable picture of himself than does one who is vaguely bounded. There was no relation between the barrier score and degree of yielding behavior. Mausner<sup>2</sup> concluded: "To summarize, this preliminary analysis tends to confirm the generalized personality description derived from the original Fisher and Cleveland work" (p. 124).

Ramer (1961) hypothesized that the high barrier person would exceed the low barrier person:

1. In initiating interpersonal communications.

2. In communicating committal, directive, and disagreeing statements rather than self-depreciating ones.

He studied 96 female subjects (in groups of four) in a setting in which each was isolated from the others. Instructions were given to write a story about several pictures and then to communicate with a fictitious partner about the story by writing messages which would presumably be delivered by the experimenter. One-third of the subjects were given no responses to the messages they sent; another third were given unfriendly replies from the fictitious partner; and the remaining third were given friendly replies. In this setting high barrier subjects sent more messages (.01 level) and more units of communication (.05 level) than the low barrier subjects. A significant difference in the same direction occurred with respect to number of committal messages as opposed to those asking for orientation. There were nonsignificant but consistent tendencies for the

barrier score to be negatively related to such variables as number of self-depreciating or passive accepting messages but positively with the number of direction giving messages. Borderline evidence was obtained that the "barrier style of behavior" was most evident under threat conditions. Ramer considered most of his hypotheses to have been confirmed or at least supported by the results.

Differences in degree of communicative expression in group situations between high and low barrier subjects has been observed by Hornstra and McPartland<sup>7</sup> to hold true also in a schizophrenic population. They report that there are significant distinctions in the ward behavior of schizophrenic patients with relatively high and low barrier scores. Patients with relatively high scores are rated by ward personnel as displaying behavior which is restless, acting out, and initiatory of interaction. This is in contrast to patients with lower barrier scores who are depicted as showing "retarded and withdrawn behavior."

The studies just summarized affirm that boundary characteristics play a meaningful part in an individual's group behavior. The more clearly articulated an individual's boundaries the greater the probability that he will seek to communicate with other group members and also that his communications will be direct and active rather than passive or self-depreciatory. With regard to the matter of suggestibility, it should be noted that the Cleveland and Morton results were supportive, and those of Mausner not supportive, of previous findings. Whether this difference is a function of the unlike methods used to measure suggestibility remains to be seen.

#### PSYCHOPATHOLOGY

The barrier and penetration scores were earlier found to discriminate schizophrenic from nonschizophrenic subjects. They did not distinguish between normals and neurotics, but roughly separated normals and neurotics (high barrier, low penetration) from schizophrenics (low barrier, high penetration) (Fisher & Cleveland, 1958b). Of course, the

<sup>7</sup> R. K. Hornstra and T. S. McPartland, unpublished progress report entitled "The Relation of Behavioral Constellations to Drug Use," 1961, National Institute of Mental Health, Grant No. MY 3308.

association of vague body boundaries with schizophrenia had for some time been remarked upon by clinical observers (e.g., Schilder, 1935).

Holtzman (Holtzman et al., 1961) applied his ink blot series to normal and pathological groups and observed that the barrier score was higher and the penetration score lower in normals than in chronic schizophrenics. However, only in the case of penetration was the difference significant (.01 level). In addition, he factor analyzed the intercorrelations of 23 indices (including barrier and penetration) derived from his ink blot test in 16 different samples of subjects. He discovered that the barrier score consistently loaded high on a factor which he associates with "well organized, ideational activity, awareness of conventional concepts" (p. 171). The penetration score loaded high on several factors related to disturbance. It was particularly identified with indicators of immaturity, bodily preoccupation, and psychopathology. These factor analytic results tie the barrier score to ego integration and the penetration score to maladjustment.

Cleveland (1960) inquired whether there would be changes in the boundary scores as schizophrenics recovered from the acute phase of their disorganization. Twenty-five male schizophrenics were evaluated upon first entering the hospital with the Lorr Multidimensional Scale for Rating Psychiatric Patients. They were evaluated again after 5 and 13 weeks of treatment involving tranquilizers. Another criterion of the patient's response to treatment was whether he attained sufficient recovery to leave the hospital. Holtzman Ink Blots were administered predrug, and 5 and 13 weeks after treatment had begun. A significant rho of .60 was found between decrement in penetration scores and decrement in the Lorr morbidity rating during the period from the onset of treatment to the fifth week. The rho for the same relationship from treatment onset to the thirteenth week was .61. There was also a significant trend (.05-.02 level) for patients judged capable of leaving the hospital to have declined in penetration. This contrasted with nondischarged patients who tended to increase their penetration scores. The barrier score failed

to be related to any of the criteria of patient change.

A second phase of this study concerned 45 schizophrenics who had been administered the Rorschach on admission and again upon leaving the hospital. Each patient was psychiatrically rated upon admission and again at time of discharge. It was established that patients rated as improved or markedly improved showed a significant decline in penetration (.01 level). For barrier the only significant change was an increase from first to second testing in the markedly improved group. Cleveland (1960) considered that his results demonstrated in the recovering schizophrenic patient a "dramatic firming up and defining of the body image boundary" (pp. 259-260).

Pankow (in Burton, 1961) observed in her psychotherapeutic work with schizophrenics that their art productions are characterized by disrupted boundaries. She specifically compared this phenomenon to the boundary indefiniteness depicted by low barrier and high penetration responses characteristic of schizophrenic populations.

Reitman (1962) studied the body image changes in neurotics and schizophrenics following sensory isolation. Twenty neurotics and 20 schizophrenics were exposed to sensory isolation conditions. Holtzman Ink Blot Tests, measures of tactile sensitivity, and estimates of body size were obtained before and after isolation. A control group received the pre- and postbattery of tests, but with isolation not intervening. No changes occurred in the scores of this group from pre- to postevaluation. However, there were significant changes in the experimental groups. The neurotics manifested decreased barrier and increased penetration scores following isolation.<sup>8</sup> The schizophrenics, contrastingly, obtained higher barrier and lower penetration scores. It was speculated that sensory isolation by minimizing stimulating input has a disruptive effect upon nonpsychotics which

<sup>8</sup> Using an entirely different methodology, Cambareri (1958) found that nonpsychotic subjects exposed to sensory isolation in a swimming tank made references to loss of body boundaries (e.g., "I had a difficult time after awhile in the tank distinguishing where I left off and the surrounding water began.").



decreases boundary definiteness. But in the case of schizophrenics the sensory isolation seemed not to be disruptive. It seemed to provide a nonthreatening pattern of stimuli which fostered reorganization and more realistic body boundaries. This latter finding was in keeping with earlier reports (Azima & Cramer, 1956; Gibby, Adams, & Carrera, 1960) concerning the therapeutic effects of isolation on schizophrenic symptomatology. It is important to note that the results from the tactile threshold and body size estimate tasks were similarly reversed for the two experimental groups and in a direction congruent with the concept that isolation produces boundary alteration. Thus, the schizophrenics showed increased tactile sensitivity and a decreased concept of body size following isolation; but the nonpsychotics showed no change in tactile sensitivity and an increased concept of body size. There are previous studies by Wapner (1958) which suggest that the pattern of changes with regard to body size estimates in the schizophrenic group are related to increased awareness of the body periphery and those in the nonpsychotic group with lessened awareness of the periphery. It should be added that in estimating the sizes of non-self objects (e.g., baseball) the two groups did not shift their judgments from test to retest. Only judgments with regard to one's own body were sensitive to the sensory isolation effects.

In the area of psychopathology the boundary scores seem to distinguish grossly between schizophrenics and nonschizophrenics.<sup>9</sup> Furthermore, there is evidence that as an individual reorganizes following schizophrenic breakdown his penetration score declines. The barrier score is less successful in

<sup>9</sup> The potential of the boundary scores for measuring other kinds of maladjustment aside from that in the category of psychosis is pointed up by Miner and De Vos (1960) finding that urban Arabs who are conflicted about their identity produce significantly higher penetration scores (.05 level) than oasis Arabs who have not yet been exposed to such extreme identity conflict. This, by the way, is confirmatory of previous data (Fisher & Cleveland, 1958b) which demonstrated that Japanese-American men who were struggling to adapt to United States life had less definite boundaries than native Japanese not beset by such identity problems.

this respect. Cleveland reported that it increased as recovery proceeded in one category of patients, but that it otherwise failed to correlate with the reorganization process. Within the context of the changes produced by sensory isolation one finds that the barrier and penetration scores vary in opposite directions in schizophrenics and nonschizophrenics. Schizophrenics react to isolation with an increase in barrier and a decrease in penetration scores, with the converse holding true for the neurotics. In the first instance isolation seems to decrease definiteness of body image boundaries and in the second it apparently helps to re-establish them.

#### DEVELOPMENTAL ASPECTS

There have been scattered indications that aspects of the developmental process (Fisher & Cleveland, 1958b; Malev, 1961) might be related to boundary attributes.

Fish (1960) scanned the boundary correlates of several developmental parameters in children. She applied a multiple-choice version<sup>10</sup> of the usual Rorschach technique to measure barrier and penetration responses in boys at ages 7 ( $N = 21$ ), 9 ( $N = 25$ ), and 11 ( $N = 25$ ). She also secured measures tapping such diverse variables as concept of time, resistance to perceiving aniseikonic induced distortions in one's mirror image, concept of one's height, and ability to define the adult role. As in a previous study (Fisher & Cleveland, 1958b), there were no indications of progressive change in boundary scores with age. It was ascertained, though, that the barrier score was positively (.05 level) correlated with ability to represent adult qualities in figure drawings at age 7. Also, the barrier score was positively correlated in 7- (.01 level) and 9-year-olds (.09 level) with a more "mature" mode of time perspective that involves the perception of future adult events as distant from, rather than close to, the present. The barrier score was not related to judgments of one's own height or to degree of distortion in self-image in-

<sup>10</sup> There was no attempt to determine the comparability of the multiple-choice method to the spontaneous response procedure usually used to obtain barrier and penetration scores.

duced by aniseikonic lenses. In general, it is interesting that despite the use of an untried method of barrier measurement, the significant results which were obtained in relation to the barrier score linked it positively with certain indices of developmental maturity.

#### DISCUSSION OF EMPIRICAL FINDINGS

The studies published since 1958 with regard to the boundary scores point up their versatility. There has been moderately good substantiation of the fact that boundary variations are accompanied by certain patterns of physiological reactivity. Both in terms of psychosomatic symptoms and measures of autonomic response the bulk of the newly reported data indicate that the higher the degree of boundary definiteness the greater the tendency to channel excitation to skin and muscle and the less the tendency to do so at interior sites like the stomach and heart. These findings have appeared in both adults and children and also in studies utilizing different designs and instrumentation. It is true that certain inconsistencies remain to be explained. For example, why does GSR sometimes correlate with barrier and sometimes not? Also, why do given autonomic measures (e.g., GSR and blood pressure) differ in the direction of their correlation with the barrier score in adults as opposed to children? Of course, the even larger task remains of clarifying the mechanisms whereby body attitudes and body reactivity patterns become linked with each other. However, what is of particular importance is that patterns of autonomic responsiveness have been demonstrated which are meaningful within a body image framework but have no apparent ties with the conventional autonomic categories (e.g., sympathetic and parasympathetic). Lacy (in Rubenstein & Parloff, 1959) has already referred to the inadequacy of the conventional categories for dealing with the observed complexity of autonomic response patterns.

The boundary studies dealing with stress tolerance and psychopathology combine to reaffirm that with increasing boundary definiteness there is a diminished likelihood of psychosis and an enhanced ability to deal

effectively with difficult, disturbing experiences (especially those involving body dis-ablement). Even the findings about small group behavior have similar connotations with regard to adjustment effectiveness; for they portray the high barrier person as coping with group interrelationships in a more independent, realistic, and yet group integrative fashion than does one who is low barrier. In addition, of course, the results from the group studies point up that the way in which an individual delineates his body limits in relation to others plays a noteworthy role in his group conduct. It is still an unsettled matter as to which aspects of group interaction are most directly related to the boundary parameter. However, the trends cited suggest that high and low barrier subjects are best distinguished in their group behavior with respect to how much each tries to communicate with others for mutual facilitation and stimulation rather than for self-centered purposes.

An unexpected fact that has emerged from the newer studies is that the boundary scores are indicators of certain kinds of change in the individual. Cleveland (1960) detected shifts in the penetration score as schizophrenics recovered from disorganization. McConnell and Daston (1961) recorded meaningful changes in the penetration scores of pregnant women from pre- to postdelivery period. Reitman (1962) discovered the exciting fact that both barrier and penetration scores are altered in neurotics and schizophrenics during sensory isolation. Furthermore, these alterations are accompanied by equivalent changes in light touch threshold and concept of one's body size. Here one sees boundary score changes correlated with such widely different phenomena as personality reorganization, the completion of pregnancy, and the impact of decreased sensory experience. Apparently, fluctuations in boundary attributes do offer some promise as indicators of certain modifications of the individual. At this point one can only conjecture whether boundary fluctuations represent initiating forces in change processes or whether they are subsidiary effects.

It should be specified that the penetration score has been the index primarily correlated



with change phenomena. The barrier score was consistently related only to changes produced by sensory isolation. This tends to be congruent with results from earlier studies (Fisher, 1959a; Fisher & Cleveland, 1958b) which indicated that the barrier score was largely a measure of persisting attitudes rather than of short-term variations in state. The penetration score, by contrast, seems to be more sensitive to immediate situational conditions.<sup>11</sup> However, this differentiation can only be considered to be a tentative one which waits further confirmation.

Actually, the relationships of the barrier and penetration scores still remain to be worked out in detail. These relationships fluctuate in different populations (although often low and tending in a negative direction). The relative usefulness of the two scores is an open question and admittedly has been neglected as a matter of inquiry. Beginning attempts to combine them into one index of boundary definiteness have been made with fair success (e.g., Fitzgerald, 1961; Malev, 1961).

In general, one seems justified in saying that the studies which are reviewed are supportive of the major past findings concerning body image boundary definiteness which have been reported.

#### INTERPRETATION OF BOUNDARY SCORES

It would be well in closing this overview to consider a general question which has been raised concerning the interpretation of the boundary scores. Wylie (1961), Mednick (1959) and others have criticized the assumption that barrier and penetration scores represent measures of body image, as such. They are inclined to view these scores as more indicative of cognitive or perceptual operations than of anything derived from

body experience. They imply that the observed network of empirical relations between the boundary scores and other phenomena could just as well be explained within a perceptual framework as via body image constructs. Basically, they question whether there is adequate support for assuming that the properties assigned to the periphery of ink blot percepts reflect how the body boundary is experienced. Such criticisms cannot be easily dismissed. The complete absence of previous empirical work with regard to body boundary feelings has made it difficult to find adequate body image criteria against which to validate the boundary properties assumed to characterize barrier and penetration responses. There are, however, certain lines of evidence which have encouraged the writer (and Sidney Cleveland) to persist in a body image conceptualization of the data which have accumulated:

1. Primary among these is the fact that the boundary scores are correlated with body phenomena to a degree unequaled by previous measures.<sup>12</sup> They predict various levels of body behavior in the way that a body image measure would be expected to do. They have shown themselves to be meaningfully related to ability to cope with disablement of one's body; to patterns of phantom sensation triggered by amputation; to experienced changes in body size and also light touch threshold following sensory isolation; to differential size judgments assigned to the right and left sides of the body; to concern about the vulnerability of the skin; and to degree of anxious concern about the body (Fisher, 1960; Fisher & Cleveland, 1958b). Incidentally, the last of these cited relationships is particularly pertinent because the boundary scores do not correlate with indices of gen-

<sup>11</sup> The sensitive response of the penetration score to variations in state is further exemplified in a report by Herron (1962). He found that when the Holtzman blots were administered under neutral conditions and under conditions designed to arouse achievement-motivation, the penetration score was the most reflective, among a variety of other indices, of the achievement condition. It was significantly lower in the achievement oriented than in the neutral state.

<sup>12</sup> In a just completed and unpublished study the writer (and Rhoda Fisher) have found a significant relationship between the barrier score and the body associations given by normal subjects. High barrier subjects gave associations which indicated that they exceeded low barrier subjects (.02 level) in the degree to which their awareness of the body boundary region (skin and muscle) exceeded their awareness of the body interior (stomach and heart). This study represents the most direct and convincing demonstration of the body image foundation of the barrier score to date.

eral anxiety which are not specifically phrased in body terms.

It is further striking that the boundary scores predict patterns of body reactivity which involve a differentiation between the boundary and nonboundary regions of the body. Whether one assumes that the scores are reflections of differential levels of activation at such body sites (inside versus outside) or actually play a role in instigating the activation differences, the fact remains that a solid correlation exists. There is force in this congruence between boundary properties assigned to the ink blot periphery and the reactivity characteristics of the body periphery versus body interior. One cannot dismiss it as mere coincidence if one considers the fact that the existence of an exterior-interior activation pattern in normal subjects was specifically deduced from the model provided by the body image boundary concept. There are few, if any, non-body-image concepts which would lend themselves to such a prediction phrased in terms of body "geography."

2. No dependable relationships have been found between the boundary scores and indices which might be considered to have cognitive or "perceptual style" connotations. The boundary scores are not consistently related to such variables as intelligence, verbal productivity, the Barron simplicity-complexity dimension, Gottschaldt figure judgments, conventional individual Rorschach determinants, speed of figure-ground alternation, rigidity, or authoritarianism.

It is recognized that more work needs to be done to establish the body image rationale which has been advanced for the boundary scores. As a matter of fact, the author has several studies under way which seek to demonstrate that an individual's introspective reports of body experience in various situations (e.g., under stress, in response to drugs, in response to placebo) conform to the particular ratio of exterior versus interior sensations that would be predicted from his barrier and penetration scores.

As a final word, it may be added that not the least of the attractions of a body image interpretation of the boundary scores is the fact that it has led to the successful testing

of a variety of novel hypotheses. Is this not a major consideration in any research strategy?

#### REFERENCES

- AZIMA, D., & CRAMER, F. J. Effects of the decrease in sensory variability on body scheme. *Canad. J. Psychiat.*, 1956, 1, 59-72.
- BARTS, G. J. The perception of body boundaries. Unpublished master's thesis, Ohio State University, 1959.
- BELOFF, J., & BELOFF, H. Perception of self and others using a stereoscope. *J. abnorm. soc. Psychol.*, 1957, 56, 87-92.
- BRODIE, C. W. The prediction of qualitative characteristics of behavior in stress situations, using test-assessed personality constructs. Unpublished doctoral dissertation, University of Illinois, 1959.
- BROWN, D. G. Psychosomatic correlates in contact dermatitis: A pilot study. *J. psychosom. Res.*, 1959, 4, 132-139.
- BURTON, A. *Psychotherapy of the psychoses*. New York: Basic Books, 1961.
- CAMBARERI, J. D. The effects of sensory isolation on suggestible and nonsuggestible psychology graduate students. Unpublished doctoral dissertation, University of Utah, 1958.
- CLEVELAND, S. E. Body image changes associated with personality reorganization. *J. consult. Psychol.*, 1960, 24, 256-261.
- CLEVELAND, S. E., & FISHER, S. A comparison of psychological characteristics and physiological reactivity in ulcer and rheumatoid arthritis groups: I. Psychological measures. *Psychosom. Med.*, 1960, 22, 283-289.
- CLEVELAND, S. E., FISHER, S., REITMAN, E. E., & ROTHBAUS, P. Perception of body size in schizophrenia. *Arch. gen. Psychiat.*, 1962, 7, 277-285.
- CLEVELAND, S. E., & JOHNSON, D. C. Personality patterns in young males with coronary diseases. *Psychosom. Med.*, in press.
- CLEVELAND, S. E., & MORTON, R. B. Group behavior and body image: A follow-up study. *Hum. Relat.*, 1962, 15, 77-85.
- DARROW, C. W. Differences in the physiological reactions to sensory and ideational stimuli. *Psychol. Bull.*, 1929, 26, 185-201.
- DASTON, P. G., & MCCONNELL, O. L. Stability of Rorschach penetration and barrier scores over time. *J. consult. Psychol.*, 1962, 26, 104.
- DAVIS, A. D. Some physiological correlates of Rorschach body-image productions. *J. abnorm. Psychol.*, 1960, 60, 432-436.
- DAVIS, R. C. Response patterns. *Trans. N. Y. Acad. Sci. Ser. 2*, 1957, 19, 731-739.
- EIGENBRODE, C. R., & SHIPMAN, W. G. The body image barrier concept. *J. abnorm. soc. Psychol.*, 1960, 60, 450-452.
- FISH, J. E. An exploration of developmental aspects of body scheme and of ideas about adulthood in grade school children. Unpublished doctoral dissertation, University of Kansas, 1960.



- FISHER, RHODA L. The effect of a disturbing situation upon the stability of various projective tests. *Psychol. Monogr.*, 1958, 72(14, Whole No. 467).
- FISHER, S. Body image boundaries in the aged. *J. Psychol.*, 1959, 48, 315-318. (a)
- FISHER, S. Prediction of body exterior vs. body interior reactivity from a body image scheme. *J. Pers.*, 1959, 27, 56-62. (b)
- FISHER, S. Right-left gradients in body image, body reactivity, and perception. *Genet. Psychol. Monogr.*, 1960, 61, 197-228.
- FISHER, S., & CLEVELAND, S. E. Body image boundaries and sexual behavior. *J. Psychol.*, 1958, 45, 207-211. (a)
- FISHER, S., & CLEVELAND, S. E. *Body image and personality*. Princeton: Van Nostrand, 1958. (b)
- FISHER, S., & CLEVELAND, S. E. A comparison of psychological characteristics and physiological reactivity in ulcer and rheumatoid arthritis groups: II. Differences in physiological reactivity. *J. psychosom. Med.*, 1960, 22, 290-293.
- FITZGERALD, W. E. A psychological factor in Legg-Calvé-Perthes disease. Unpublished doctoral dissertation, Harvard University, 1961.
- GIBBY, R. G., ADAMS, H. B., & CARRERA, R. N. Therapeutic changes in psychiatric patients following partial sensory deprivation. *Arch. gen. Psychiat.*, 1960, 3, 33-42.
- HERRON, E. W. Intellectual achievement-motivation: A study in construct clarification. Unpublished doctoral dissertation, University of Texas, 1962.
- HOLTZMAN, W. H., THORPE, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality*. Austin: Univer. Texas Press, 1961.
- LANDAU, M. F. Body image in paraplegia as a variable in adjustment to physical handicap. Unpublished doctoral dissertation, Columbia University, 1960.
- LHERMITTE, J. *L'image de notre corps*. Paris: Editions, de la Nouvelle Revue Critique, 1935.
- McCONNELL, O. L., & DASTON, P. G. Body image changes in pregnancy. *J. proj. Tech.*, 1961, 25, 451-456.
- MALEV, J. S. Body image and physiological reactivity in children. Unpublished master's thesis, Baylor University, 1961.
- MEDNICK, S. The body's barriers go Rorschach. *Contemp. Psychol.*, 1959, 4, 276-277.
- MENER, H. M., & DE VOÏS, G. *Oasis and Casbah: Algerian culture and personality in change*. Ann Arbor: Univer. Michigan Press, 1960.
- RAMER, J. C. The Rorschach barrier score and social behavior. Unpublished doctoral dissertation, University of Washington, 1961.
- REITMAN, E. E. Changes in body image following sensory deprivation in schizophrenic and control groups. Unpublished doctoral dissertation, University of Houston, 1962.
- RUBINSTEIN, E. A., & PARLOFF, M. B. *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.
- SCHILDER, P. *The image and appearance of the human body*. London: Kegan Paul, Trench & Trubner, 1935.
- SECORD, P. F. Objectification of word association procedures by the use of homonyms: A measure of body cathexis. *J. Pers.*, 1953, 21, 479-495.
- WAPNER, S., WERNER, H., & COMALLI, P. E. Effect of enhancement of head boundary on head size and shape. *Percept. mot. Skills*, 1958, 8, 319-325.
- WARE, K. S., FISHER, S., & CLEVELAND, S. E. Body image boundaries and adjustment to poliomyelitis. *J. abnorm. soc. Psychol.*, 1957, 55, 88-93.
- WILLIAMS, R. L. The relationship of body image to some physiological reactivity patterns in psychosomatic patients. Unpublished doctoral dissertation, Washington University, 1962.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.
- WOLFF, W. *The expression of personality*. New York: Harper, 1943.
- WYLIE, R. C. *The self concept*. Lincoln: Univer. Nebraska Press, 1961.

(Received June 7, 1962)

## SYMBOLIC LEARNING AND READING RETARDATION

RICHARD H. WALTERS

AND

IRENE KOSOWSKI<sup>1</sup>

*University of Toronto*

*Toronto Psychiatric Hospital*

It was predicted that advanced, average, and retarded readers would also be advanced, average, and retarded in both the number of trials required to master symbolic learning tasks and in their response latencies to visual and auditory symbols. Some confirmation of the prediction was obtained, but both the provision of a reward and the effect of practice increased the efficiency of retarded readers, especially on the auditory task. The data were interpreted as suggesting that retarded readers need an incentive in relatively difficult learning situations and also that, unless highly motivated, retarded readers tend to be less attentive to stimuli. It was further suggested that difficulties in symbolic learning may be, in part, a function of reduced ability to attend to stimuli.

Walters and Doan (1962) suggested that retarded readers have difficulty in learning to associate symbols with responses and that this difficulty is not confined to tasks in which the symbol is a printed word. This suggestion was supported by a study of symbolic learning in advanced, average, and retarded readers under reward and nonreward conditions. The subjects were Grade 7 and 8 boys from two public elementary schools. Boys in one school were assigned to the reward condition; children in the second school were not rewarded. A multiple-choice apparatus with four compartments, each opened by a separate door, was used. The subject was asked to find a neutral stimulus object hidden in one of the compartments, each of which was symbolized by a light of a particular color. His task was to learn the associations between the colors and the compartments. Under reward conditions retarded readers performed significantly less efficiently than did either advanced or average readers. Provision of a reward produced quite marked improvement for retarded readers, but the performance of the other two groups was influenced very little by this procedure. The authors concluded that there are indications that the performance of retarded readers, particularly in the area of symbolic learning, can be improved by the provision of

a reward; however, they suggested that their results might have been more conclusive had the reward been more elaborate and more appropriate for the subjects' age levels.

Since Walters and Doan were interested in several aspects of the perceptual and cognitive functioning of retarded readers, only a part of their study was devoted to symbolic learning. Their use of only a visual stimulus limits the scope of any generalization that may be made on the basis of their study. Moreover, the use of students from two different schools in the reward and nonreward conditions, and the consequent possibility of confounding the influence of reward with influences of different teaching methods, socioeconomic status, and other school-related factors reduce confidence in their findings relating to the differential effects of motivational level on the three groups of readers.

Walters and Doan used only a comprehension test as an index of reading proficiency. Reading comprehension correlates highly with verbal intelligence test scores (Hage & Stroud, 1959), and a composite reading score is a more valid index of reading proficiency than any one test alone (Reed & Pepper, 1957). Consequently Walters and Doan may not have adequately assessed the discrepancy between the reading proficiency scores and the intelligence test scores of their children, i.e., the measure on which their division of pupils into different reading groups was based.

The present study represents an attempt to extend and refine the procedures used by

<sup>1</sup> The authors wish to express their appreciation to the Scarborough Board of Education and the Principal and Staff of Clairlea Public School for their cooperation in this study, and to Christine Orbell for assistance in collecting the orthorater and audiometer data.



Walters and Doan in investigating symbolic learning in retarded readers. It differs from the earlier study in the following respects:

1. The inclusion of two types of stimuli (symbols)—auditory and visual.

2. The use of a composite reading score based on tests of reading speed, comprehension, and vocabulary.

3. The selection of rewards only after consultation with school authorities to determine their appropriateness to the children's age levels.

4. The use of children from only one school which was located in a socioeconomically homogeneous suburban community.

The design also permitted an investigation of transfer effects from one sense modality to another and consequently provided data relevant to the hypothesis advanced by Raab, Deutsch, and Freedman (1960) that retarded readers cannot shift responses as easily as good readers. More directly, it provided evidence concerning the extent to which children at different reading levels benefit from practice in symbolic learning.

There is some advantage, at least in preliminary investigations of reading retardation, in confining attention to children receiving education through the regular school curriculum, since placement in special classes or selection for remedial reading groups may produce emotional reactions which could influence performance in selected perceptual and cognitive tasks.

Reading disability is much more prevalent among boys than among girls. For this reason, and because the motivational patterns of girls in social learning tasks differ from those of boys (Crandall, 1963), only boys were used as the subjects in this study.

Studies of development of perceptual skills in children indicate that performance on perceptual tasks improves considerably with age during early and middle childhood (Bender, 1949; Coleman, 1953). In view of this finding, boys under 11 years of age were excluded from the present experiment.

In the present study the following hypotheses were tested:

1. In a multiple-choice task involving symbolic learning, retarded readers perform more poorly than average readers, and average

readers perform more poorly than advanced readers.

2. Retarded readers give slower reaction times to both auditory and visual stimuli in a multiple-choice task than average readers, and average readers give slower reaction times than advanced readers.

3. Provision of an incentive reduces differences in performance among the three groups of readers.

## METHOD

### *Subjects*

The subjects were drawn from Grades 6, 7, and 8 boys attending a single suburban public school.

The selection procedures which led to the classification of the subjects into advanced, average, and retarded reading groups were similar to those described by Walters and Doan (1962). They involved a comparison of each subject's relative standing in reading (average reading grade) and his relative standing in intelligence.

The Dominion Group Test of Learning Capacity and Gates' Reading Survey were administered to all boys in Grades 6, 7, and 8. Grades 6 and 7 tests were scored by the experimenter, and Grade 8 tests by the school principal.

The raw scores for each test were transformed into  $z$  scores in order to estimate the discrepancy between each child's reading level and his intelligence. The  $z$  scores were computed separately for each test in each grade in order to restrict the comparison of each pupil's performance to that of his peers in age and educational level. Boys whose  $z$  scores in reading minus their  $z$  scores in intelligence were greater than  $+1.5$ , were classified as advanced readers. Those with a discrepancy score greater than  $-1.5$  were considered retarded readers, and those with discrepancy scores between  $+1.39$  and  $-1.39$  constituted the average reader group.

The IQ range of subjects used in this study was restricted to between 90 and 126 in view of the difficulty in interpreting discrepancies between intelligence and achievement in children at extreme ranges of ability.

Boys less than 11 years old were excluded from study. Boys with known visual anomalies and hearing loss (as measured by a test of auditory acuity) were eliminated from the sample, as were those who had not been educated entirely in Canada. Also excluded were boys reported by school authorities to have emotional or behavior problems.

Seventy-two boys—24 advanced, 24 average, and 24 retarded readers—who met all the above requirements were selected. Their age range was from 11.0 to 15.8 years, with a mean IQ of 106 for Grade 6, 107 for Grade 7, and 104 for Grade 8. The mean IQ for all subjects was 105.

Advanced, average, and retarded readers were assigned to reward and nonreward conditions in such

a way as to maintain, as far as possible, a balance among the six subgroups in age, grade level, and in reading-intelligence discrepancy scores.

### Apparatus

The apparatus consisted of a multiple-choice box previously described by Walters and Doan (1962), and the procedures used for presenting visual symbols were precisely the same as in the earlier study. Each compartment was randomly paired with lights of a particular color and the subject's task was to learn the correct color-compartment associations.

For the administration of the auditory task the visual mechanism and buzzer were disconnected. A pure-tone generator with earphones, attached to the multiple-choice box, was used to present auditory symbols. When the box was completely closed, a tone, selected by means of a manual presetting behind the box, became audible through the earphones. Each of the four compartments of the box was coupled at random with one distinct tone. The frequencies used were approximately 100, 175, 420, and 2,000 cps; these frequencies were selected as being readily distinguishable one from another.

The apparatus enables the latencies of the subjects' responses on each trial to be automatically recorded (Walters & Doan, 1962).

### Procedure

Each subject was conducted by the experimenter to a small room, which was illuminated by artificial light and contained only a desk and chairs in addition to the test equipment. The subject was asked to stand in front of the multiple-choice box and was then introduced to the test procedures.

The instructions for the administration of both the visual and auditory tasks varied according to the order in which these tasks were presented. If the visual task was presented first, the experimenter gave the following instructions:

I am going to hide this nebbish (holding up the test object) in one of these four boxes. (Demonstrate.) I want you to see if you can find it. When you hear a buzzer, watch these lights. (Demonstrate.) As soon as you see a light go on, open the door of the box in which you think the nebbish is hidden. If you watch the lights carefully, you can learn to choose the correct box every time.

The experimenter seated herself behind the box and proceeded to place the nebbish in the back of the various compartments in a predetermined random order. On each trial, she pressed the appropriate light button, which also served to activate the buzzer. The experiment was continued until the subject had completed eight successive correct trials. The latency of the subject's response and his choice of compartment were recorded by the experimenter for each trial.

In order to separate in time the administration of the two symbolic learning tasks, a coding task,

taking approximately 15 minutes, was introduced. This task was chosen for exploratory purposes only.

During the administration of the auditory task, the subject, wearing earphones, again stood in front of the multiple-choice box. Following a demonstration of the four tones, he was instructed as follows:

Now I am going to hide this nebbish again in one of the four compartments of this box. This time I want you to listen to the tones in your earphones. As soon as you hear a tone, open the door of the box in which you think the nebbish is hidden. If you listen carefully, you can learn to choose the correct box every time.

The latency of the subject's responses and his choice of compartment were recorded for each trial. The trials were repeated until eight successive correct responses were obtained.

When the auditory task was presented first and the visual task second, instructions for each task were appropriately modified.

When all other testing was completed, each child was handed a color naming chart (Wells & Ruesh, 1945) and asked to name colors in quick succession from left to right, in order to ensure that poor performance was not due to difficulty in identifying colors. No subject included in the study had a color-identification problem.

Each subject was asked to refrain from discussing the experiment with other children until it was completed.

*Order of presentation.* The order of testing of the subjects was determined by the alphabetical order of their names. One half of the children in each of the six subgroups was given the auditory task first, while the remaining children were given the visual task first. Assignment to one or other order of presentation was made at random.

Since knowledge that other subjects were being rewarded for participation in the study could have influenced the attitudes and performance of the subjects in the nonreward group, all nonreward subjects were tested before testing of reward subjects began. An interval of approximately 3 weeks, which included the Christmas vacation, separated the testing of nonreward and reward subjects.

*Rewards.* Reward subjects were shown an array of games and attractive objects (i.e., miniature chess sets, pens, key chains, stamps, etc.). They were allowed to examine the display for a few minutes and then, before commencing the first task, were told: "If you try hard and do well on these tasks which I will ask you to do, you can have one of these prizes. Remember, you can get the prize only if you do well on these tasks."

At the conclusion of testing each subject was told that he had done well and that he would receive the object of his choice when the experimenter had finished her work at the school. In an attempt to minimize communication between the subjects and to reduce the children's curiosity concerning the nature of the experiment, distribution of prizes were withheld until the conclusion of the experiment.



Following the completion of testing, all subjects (reward and nonreward) who took part in the project were given a "prize" for their participation.

*Supplementary tests.* Measures of the subjects' binocular visual acuity were obtained on a Bausch and Lomb modified orthorator. Auditory acuity was tested with a Maico audiometer. These measures provided a check on the accuracy of school reports concerning absence of visual or auditory defects. There was no evidence of differences in visual or auditory acuity among the three reading groups.

### RESULTS

Two measures of each subject's performance were secured from the symbolic learning tasks: number of trials to criterion and median reaction time of the subject's responses. Medians were chosen in preference to mean reaction times since for many subjects the initial response latency was disproportionately slow.

#### *Analysis of Trials-to-Criterion Measure*

Table 1 shows the mean number of trials to reach criterion for all subgroups of subjects on both the visual and auditory tasks. An analysis of variance, the results of which are given in Table 2, indicated that there were differences among the groups significant at the .05 level, and a Tasks  $\times$  Orders effect sig-

TABLE 1  
MEAN NUMBER OF TRIALS TO CRITERION ON THE  
SYMBOLIC LEARNING TASKS

	Visual		Auditory	
	Visual first	Auditory second	Visual second	Auditory first
Advanced readers				
Rewarded	93.67	66.17	35.50	69.33
Not rewarded	71.67	47.17	25.17	61.67
Average readers				
Rewarded	60.67	41.83	21.83	102.00
Not rewarded	97.67	37.66	26.33	77.33
Retarded readers				
Rewarded	95.83	48.33	45.67	81.83
Not rewarded	102.17	41.50	30.83	144.67

Note.— $N = 6$  in each subgroup.

nificant at the .001 level. This latter finding may reflect a kind of transfer effect, i.e., subjects with previous training on the visual task did better on the auditory task than subjects without previous training, and vice versa. Subsequent *t* tests of group mean differences showed that retarded readers performed more poorly than either advanced or average readers (differences being significant at the .01 level), but that advanced and average readers did not differ.

In addition to the overall analysis of vari-

TABLE 2  
ANALYSIS OF VARIANCE OF NUMBER OF TRIALS TO CRITERION  
ON THE SYMBOLIC LEARNING TASK

Source	SS	df	MS	F
Reading groups (G)	7,573.875	2	3,786.938	3.38*
Reward conditions (C)	0.340	1	0.340	<1
Order of presentation (O)	1,687.840	1	1,687.840	1.51
G $\times$ C	4,422.931	2	2,211.466	1.98
G $\times$ O	4,241.514	2	2,120.757	1.90
C $\times$ O	85.563	1	85.563	<1
G $\times$ C $\times$ O	4,182.123	2	2,091.062	1.87
Error(b)	67,147.750	60	1,119.129	
Tasks (T)	3,164.063	1	3,164.063	2.21
T $\times$ G	454.318	2	227.159	<1
T $\times$ O	87,172.562	1	87,172.562	60.82***
T $\times$ C	0.007	1	0.007	<1
T $\times$ G $\times$ O	7,114.348	2	3,557.174	2.48
T $\times$ G $\times$ C	6,854.652	2	3,427.326	2.39
T $\times$ C $\times$ O	2,626.562	1	1,313.281	<1
T $\times$ G $\times$ C $\times$ O	3,669.239	2	1,833.115	1.28
Error(w)	85,990.749	60	1,433.179	

\*  $p < .05$ .  
\*\*\*  $p < .001$ .

TABLE 3  
ANALYSIS OF VARIANCE OF TRIALS TO CRITERION ON VISUAL TASK ONLY

Source	SS	df	MS	F
Groups (G)	3,678.250	2	1,839.125	1.20
Order of presentation (O)	56,560.056	1	56,560.056	36.74***
Reward conditions (C)	0.222	1	0.222	<1
G × C	4,259.695	2	2,129.848	1.39
G × O	221.027	2	110.51	<1
G × C × O	1,578.583	2	789.291	<1
Error(b)	92,093.666	60	1,534.894	

\*\*\*  $p < .001$ .

ance, separate analyses of variance were carried out for the visual and auditory tasks. In neither case was there a significant main effect related to reading level. For the visual task (Table 3), the subjects who had had pretraining on the auditory task performed better than the subjects who had had no pretraining, but no other effect was significant. On the auditory task (Table 4), there was a significant Groups × Conditions interaction ( $p < .05$ ). Under nonreward conditions both advanced and average readers performed better than retarded readers, differences being significant at the .01 and .02 levels, respectively, but average and advanced readers did not differ. Retarded readers under the reward condition performed significantly better than retarded readers under the nonreward condition ( $p < .05$ ); the performance of average and advanced readers was not significantly influenced by reward. In addition, the effect of previous training on the visual task was

not uniform for the three groups of readers, as indicated by a Groups × Orders interaction significant at the .01 level. In the case of average and retarded readers, the groups pre-trained on the visual task performed better on the auditory task than the corresponding groups without pretraining ( $p < .01$  and  $p < .001$ , respectively).

#### Analysis of Median Reaction Times

The median auditory and visual reaction times for each subject were estimated, and these medians were then averaged to give the group means shown in Table 5.

Table 6 shows the results of an analysis of variance of subject's median reaction times for the multiple-choice tasks. There was a significant Groups × Orders effect, reflecting the fact that when the auditory task was given first, retarded readers responded more slowly than either advanced or average readers ( $p < .001$  in each case), whereas

TABLE 4

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION ON AUDITORY TASK ONLY

Source	SS	df	MS	F
Groups	4,350.250	2	2,175.125	2.10
Order	32,300.347	1	32,300.347	31.24***
Conditions	0.125	1	0.125	<1
G × C	7,017.583	2	3,508.791	3.39*
G × O	11,134.528	2	5,567.264	5.38**
C × O	1,830.125	1	1,830.125	1.77
G × C × O	6,273.083	2	3,136.542	3.03
Error(b)	62,044.833	60	1,034.081	

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

TABLE 5  
GROUP MEANS OF MEDIAN REACTION TIMES IN THE AUDITORY AND VISUAL TASKS

	Visual		Auditory	
	Visual first	Auditory second	Visual second	Auditory first
Advanced readers				
Rewarded	190.50	173.67	155.00	191.67
Not rewarded	149.83	177.50	209.83	231.17
Average readers				
Rewarded	203.50	192.50	157.33	170.17
Not rewarded	201.83	176.17	192.50	220.83
Retarded readers				
Rewarded	166.67	174.33	239.33	321.17
Not rewarded	165.00	178.83	220.33	295.67

Note.—1/100-sec. units.



TABLE 6

ANALYSIS OF VARIANCE OF MEDIAN REACTION TIMES

Source	SS	df	MS	F
Groups	35,413.097	2	17,706.548	2.13
Conditions	1,750.028	1	1,750.028	<1
Orders	51,680.444	1	51,680.444	6.23**
G × C	5,482.764	2	2,741.138	<1
G × O	71,132.681	2	35,566.340	4.29*
C × O	8,804.695	1	8,804.695	<1
G × O × C	13,865.180	2	6,932.590	<1
Error(b)	497,857.333	60	8,297.722	
Tasks	15,876.000	1	15,876.000	8.47**
T × G	11,634.042	2	5,817.021	3.10
T × O	16,986.778	1	16,986.778	9.06**
T × C	220.028	1	220.028	<1
T × G × C	418.597	2	209.298	<1
T × G × O	3,029.346	2	1,514.673	<1
T × O × C	448.027	1	448.027	<1
T × C × G × O	3,309.849	2	1,654.925	<1
Error(w)	112,454.333	60	1,874.239	

\*  $p < .05$ .\*\*  $p < .01$ .

there were no differences among advanced, average, and retarded readers who were given the visual task before the auditory task.

Visual and auditory reaction times were also analyzed separately. No significant differences emerged from the analysis of visual reaction times. The analysis of auditory reaction times (Table 7) yielded a Groups × Pretraining interaction significant at the .05 level. Retarded readers who had had no pretraining on the visual task responded significantly more slowly than either advanced or average readers with no pretraining ( $p < .01$ ); there were no differences among advanced, average, and retarded readers who had received prior training.

TABLE 7

ANALYSIS OF VARIANCE OF MEDIAN LATENCIES FOR AUDITORY TASK ONLY

Source	SS	df	MS	F
Groups	41,430.778	2	20,718.890	4.00*
Conditions	1,605.556	1	1,605.556	<1
Orders	63,962.722	1	63,962.722	12.34***
G × C	3,640.777	2	1,820.389	<1
G × O	48,769.777	2	24,384.889	4.70*
C × O	2,640.222	1	2,640.222	<1
G × C × O	7,531.445	2	3,675.723	<1
Error(b)	310,935.333	60	5,182.256	

\*  $p < .05$ .\*\*\*  $p < .001$ .

TABLE 8

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION ON SYMBOLIC LEARNING TASKS FOR NONREWARD SUBJECTS: FIRST ADMINISTRATION ONLY

Source	SS	df	MS	F
Groups	19,778.389	2	9,889.195	5.78**
Tasks	148.028	1	148.028	<1
G × T	6,811.055	2	3,405.528	1.99
Error(b)	51,337.500	30	1,711.250	

\*\*  $p < .01$ .*Supplementary Analyses*

Since both pretraining and the introduction of reward appeared to modify group differences, further analysis were carried out in which only the data for the first task performed by nonrewarded subjects were included.

Table 8 shows the results of an analysis of variance of trials to criterion for the three groups of subjects on the task, auditory or visual, which was first administered to them. Differences among the three reading groups were significant beyond the .01 level. Since there was no significant difference between tasks and no interaction effect, data for the auditory and visual tasks were combined for subsequent  $t$  tests of differences between pairs of reading groups. These tests indicated that retarded readers had performed more poorly than either advanced or average readers ( $p < .01$  and  $p < .05$ , respectively). No difference was found between the performance of advanced and average readers.

A corresponding analysis of variance of median latencies is shown in Table 9. A difference between tasks significant at the .01

TABLE 9

ANALYSIS OF VARIANCE OF MEDIAN REACTION TIMES FOR SYMBOLIC LEARNING TASKS FOR NONREWARD SUBJECTS: FIRST ADMINISTRATION ONLY

Source	SS	df	MS	F
Groups	9,526.889	2	4,763.444	<1
Tasks	53,361.000	1	53,361.000	6.72**
G × T	18,788.667	2	9,394.334	1.18
Error(b)	238,318.667	30	7,943.956	

\*\*  $p < .01$ .

level reflects the fact that the children in general responded more slowly to the auditory task than to the visual task. Both the differences among the reading groups and the interaction effect failed to reach an acceptable level of significance.

### DISCUSSION

The study generally confirms the earlier finding of Walters and Doan (1962) that retarded readers have difficulties with symbolic learning. However, significant interactions involving the effects of reward and pretraining do not permit an unconditional statement about differences among the three groups of readers.

The provision of an additional incentive in the form of a reward did not equally improve the performance of all subjects. Neither the performance of advanced nor of average readers was significantly affected by reward; on the other hand, the performance of retarded readers who were rewarded was, on the auditory task, significantly better than that of retarded readers who were not rewarded. In fact, retarded readers who were rewarded on this task did as well as children in the other reading groups. The auditory task appears to have been the more difficult, since it took longer for all groups to learn than did the visual task. It may be that retarded readers, when a learning task is relatively difficult, need an additional incentive besides that inherent in the solution of the problem itself. On the other hand, auditory discrimination may well require more effort of attention on the part of the subject and it is possible that retarded readers, unless highly motivated, are less attentive to stimuli.

The highly significant transfer effect from the first to the second task suggests that once the general nature of the symbolic learning problem is understood, the principles involved can be generalized to somewhat similar situations to facilitate the learning process. This transfer effect was most noticeable in the case of retarded readers, who performed as well as the other two groups of readers on the second of the tasks administered, whether visual or auditory. Moreover, retarded readers who had had pretraining on the visual task responded significantly more quickly to

the auditory symbols than those who had had pretraining.

On the basis of previous studies it was predicted that retarded readers would have slowest reaction times. This hypothesis was supported in the case of the auditory task and particularly when this task was administered without pretraining. On the visual task no significant differences were found among the response latencies of the three reading groups. This finding is in direct contrast with the findings of earlier studies by Raab et al. (1960) and Walters and Doan (1962), in which it was found that retarded readers were significantly slower than more advanced readers in responding to visual stimuli. This contradiction may well be a function of the differences in the nature of the tasks employed in the three studies. Both Raab et al. and Walters and Doan used simple reaction-time tasks, whereas the symbolic learning tasks involved the additional elements of a multiple-choice situation.

The findings of the present study thus suggest that in a complex task the speed of learning of retarded readers is more likely to vary with practice and motivational level than that of advanced or average readers. Moreover, reaction times of retarded readers are also susceptible to the influence of practice. Thus, it seems that children who are average or better in reading perform in a more stable manner, from condition to condition, than do children whose reading achievement is poor.

Since the effect of pretraining was greater for retarded readers than for the other two groups of children, it seems possible that, in addition to the fluctuations in motivation often associated with underachievement, other factors contribute to the learning difficulties of retarded readers. A deficit in attention may be one of these. In a recent paper, Gardner and Lohrenz (1961) suggest that individual differences in assimilation are in part a function of characteristic differences in the degree to which individuals attend to stimuli. It is not inconceivable that the difficulties which retarded readers demonstrate in various types of associative and symbolic learning tasks are not primarily due to a deficit in their ability to form associations, but rather are a function of reduced ability to attend to stimuli. In



view of this, it may be advisable to investigate the role of attention in associative learning in relation to reading achievement.

#### REFERENCES

- BENDER, LAURETTA. Psychological principles of the Visual-Motor Gestalt Test. *Trans. N. Y. Acad. Sci.*, 1949, 11, 164-170.
- COLEMAN, J. C. Perceptual retardation in reading disability cases. *J. educ. Psychol.*, 1953, 44, 497-503.
- CRANDALL, V. J. Achievement. *Yearbk. Nat. Soc. Stud. Educ.*, 1963, in press.
- GARDNER, R. W., & LOHRENZ, L. J. Attention and assimilation. *Amer. J. Psychol.*, 1961, 74, 607-611.
- HAGE, D. S., & STROUD, J. B. Reading proficiency and intelligence scores, verbal and nonverbal. *J. educ. Res.*, 1959, 52, 258-262.
- RAAB, SHIRLEY, DEUTSCH, M., & FREEDMAN, A. M. Perceptual shifting and set in normal school children of different reading achievement levels. *Percept. mot. Skills*, 1960, 10, 187-192.
- REED, J. C., & PEPPER, R. S. The interrelationship of vocabulary, comprehension, and rate among disabled readers. *J. exp. Educ.*, 1957, 25, 331-337.
- WALTERS, R. H., & DOAN, HELEN. Perceptual and cognitive functioning in retarded readers. *J. consult. Psychol.*, 1962, 26, 355-361.
- WELLS, F. L., & RUESH, J. *Mental examiner's handbook*. New York: Psychological Corporation, 1945.

(Received June 11, 1962)

## THE USE OF AN OPEN-ENDED GROUP IN THE INTAKE PROCEDURE OF A MENTAL HYGIENE UNIT<sup>1</sup>

ANDREW S. DIBNER, ROBERT D. PALMER,<sup>2</sup> BENJAMIN COHEN, AND ARNOLD G. GOFSTEIN

*Veterans Administration Mental Hygiene Clinic, Boston, Massachusetts*

Several clinics have found screening in small groups on Intake to be beneficial in reducing waiting lists and evaluating patients for treatment. Others have used small short-term groups to orient and educate patients to treatment. We have run a continuous, open-ended Intake group for these purposes and to provide nuclei for other closed therapy groups. The result has been elimination of the waiting list and demonstrated improvement in patient preparation for psychotherapy. During a period of 18 months, the 61 patients referred to the open-ended group stayed significantly longer in treatment than a control group. A replication using another therapist yielded a similar result.

As is the case with many large clinics, the Veterans Administration Mental Hygiene Clinic in Boston was troubled with the problem of a growing waiting list of patients seeking psychotherapy. The usual Intake procedure of assigning patients deemed acceptable for psychotherapy to therapists on the basis of openings as they occurred was slow, often resulting in a pile-up of cases. An experimental group therapy procedure, designed to screen candidates for psychotherapy and to function continuously as an open-ended temporary treatment group, was instituted to make a waiting list unnecessary.

The special group was expected to help in three areas—first, reduce or eliminate the waiting list; second, improve proper placement in treatment; third, better prepare patients to accept psychotherapy.

Many clinics have reported waiting lists from 6 months to 2 years. This seems to be especially a problem with child guidance clinics. In some the wait is between the first contact and Intake evaluation; in some (as in our clinic) it is between Intake evaluation and first therapy appointment. At the time our study was initially started in 1958, we had about 60 people on our waiting list. Waiting proves to be an inconvenience and

frustration to clients who are seeking help. Waiting has been claimed by many to augment ailments, to make resistance to therapy more obdurate, and to prolong the therapeutic process. For some patients a disorder may have become chronic by the time he is seen for the initial interview. However, it should also be mentioned that in many cases a waiting list serves as a benefit because it gives some patients time to solve their problem on their own. It often gives time for external situations to develop, thus relieving pressure, and in the cases of children who need help, it gives the child a chance to mature. In general, however, most agree that any technique which would reduce a waiting list would be beneficial.

The second problem area of Intake at which our open-ended group technique was directed was that of screening for psychotherapy. Could this new procedure provide a fresh approach to social diagnosis and augment the impressions gained through interviews by the social worker and psychiatrist? Could it shed light on prognosis and indicate the patient's capacities to benefit from individual psychotherapy or group psychotherapy?

The third Intake problem of concern is that of preparing the patient to accept psychotherapy. The high dropout rate in psychotherapy after being screened for acceptability is evidence of patients' unpreparedness. Many investigators have complained of the waste of professional effort and time that this causes:

<sup>1</sup> A portion of this paper was read at the Eastern Psychological Association Meeting, Atlantic City, April 1961.

<sup>2</sup> Now at Worcester State Hospital, Worcester, Massachusetts.



the hours of psychiatrists', psychologists', and social workers' time that go into Intake and the further hours that are scheduled for patient appointments where the patient does not show up. In our clinic, where only about 50% of the applicants are offered psychotherapy, 45% drop out by the fifth session. Kurland (1956) reports similar statistics at other Veterans Administration clinics.

Various methods have been sought to end this waste, to better select and to motivate patients for psychotherapy. For example, Bitterman (1958) reports using a program of telephone evaluations to judge applicants' tolerance for waiting, and scheduling first appointments appropriately.

Taboroff, Brown, Kormer, Reiser, Talmadge, Goates, and Stein (1956); Stone, Parloff, and Frank (1954); and Peck (1953) report doing Intake diagnosis in short-term groups. Taboroff found that diagnostic ratings derived from observations of behavior in groups were similar to those accomplished by psychiatric interviews and psychological tests. Peck found that patients could be oriented to psychotherapy while they were being evaluated for psychotherapy.

Other types of short-term screening groups were described by Ginott (1956) and by Hotkins, Kriegsfeld, and Sands (1958). These were for parents who applied to child guidance centers. It was found that appropriate referrals either within the agency or elsewhere in the community could be rapidly made and that the early dropouts could also be quickly and economically identified (Ginott & Blek, 1959).

#### METHOD

##### *Subjects*

The experimental group (E) consisted of 61 male veteran patients assigned to the open-ended therapy group by Intake psychiatrists during a 15-month period. Although the group was intended to exclude patients diagnosed as psychotic, some were accidentally referred. The control group (C) was chosen to represent a typical cross-section of the usual Intake referral (directly to a therapist) during the time the experimental group was operating. Two control subjects were selected for each experimental subject, one chosen from the Intake list of the day before the experimental case was referred and one from the day after. Thus a control group of 122 cases was formed. The average age of patients in both groups was approximately 39 years.

Sixty percent of the cases in each group had diagnoses of psychoneurosis, the rest formed approximately equal groups of personality disorders, psychophysiological disorders, or psychoses, borderline or in remission.

At the time of the study all patients seeking treatment were seen that day by a social worker and psychiatrist. The psychiatrist would decide if psychotherapy would be offered; then the patient would be assigned to an opening in someone's schedule or would be put on a waiting list. The first batch of patients in the open-ended group came from the waiting list; subsequently they came from Intake.

##### *Open Group Therapy Procedures*

The therapist had a brief individual interview with each new patient before his entry into the ongoing group. He described the clinic's treatment facilities and presented the open-ended group to the patient as a "trial" situation which would be used by the patient and the therapist to decide which type of treatment, if any, was most appropriate for him. The patient was told that, after a few meetings of the group, he and the therapist would meet again to discuss their decision as to further treatment. The treatment proper involved weekly meetings, during which the therapist fostered group awareness and discussion of their experience and feelings in relation to symptoms. The therapeutic technique differed from that used in the usual "closed" group at our clinic in that the therapist was more active in introducing new members and in orienting his remarks toward describing what group therapy is about. The therapist expressed clarification as simply as possible. He would also occasionally arrange individual meetings with patients after a group meeting to discuss with them their behavior in group and their reactions to the treatment.

The group roster ranged from 4 to 12 and there would be an average of 6 to 7 at meetings. When the therapist felt comfortable with a decision as to what further treatment would be best for the individual case, he would seek out the proper opening and reassign the patient after first discussing it with him after a group meeting. When transferring to group therapy, he would always transfer patients in pairs or other groupings. On three occasions he transferred five or more at a time as a nucleus of a new group. Patients stayed in the group from 1 to 30 sessions, averaging 6.2 meetings.

##### *Procedures for Evaluating the Treatment*

In order to evaluate the effectiveness of the open-ended group, attendance and discharge records were reviewed to determine the length of stay in treatment for patients in E and C groups. Eight patients in the E group, (13%), and 13 patients in the C group, (11%), did not report for even one therapy session. Since these patients could not be said to have experienced any treatment at all, they were dropped from further study. All further com-

parisons were done with patients who attended at least one therapy session. An arbitrary cutoff date, 3 months after the last patient was referred by Intake, was chosen, and then the number of sessions attended by each patient from date of referral to either date of discharge or cutoff date was counted.

### RESULTS

Patients in the Experimental group stayed an average of 6.2 meetings in the open-ended group and then went on for an average of 15.5 further sessions in other treatment by the cutoff date, totaling 21.7 therapy meetings. During this period, patients in the Control group stayed an average of only 13.9 meetings. The means are significantly different by *t* test.

Another method of comparison of the groups is illustrated in Table 1. This shows that a greater percentage of E group than C group remains in treatment over time.

In order to check if these differences were due to factors other than the experimental variable, the patients were compared on relevant dimensions.

The patients in E and C groups were found not to differ significantly on age, number of previous starts in psychotherapy, year of first seeking psychotherapy, average previous length of stay in psychotherapy, and total accumulated months of previous psychotherapy. Thus, previous experience with psychotherapy could not have produced our results. The two groups were different in diagnosis in that a higher proportion of patients in the E group were diagnosed Personality Disorder and a higher proportion of patients in the C group were diagnosed

Psychotic. However, since patients with diagnoses of personality disorder had a lower attendance record than the group average, and patients with psychotic diagnoses had a higher attendance record than the average, the group differences on these dimensions could not have accounted for the experimental findings. Rather they work in the reverse direction.

### DISCUSSION

First of all one might ask, "Does this sort of procedure have any therapeutic effectiveness, or is it wholly diagnostic and educational?" Although this question was not formally explored, it is the opinion of those consultants involved in the experiment that it was therapeutic. Three patients were discharged as improved directly from the open-ended group, and after several meetings many other patients verbalized lessening of anxiety feelings or reduction in physical complaints. From the standpoints of content, activity of interaction, and extent of patient participation tape recordings of most sessions of the group would be indistinguishable from recordings of other "closed" therapy group meetings. In most ways the therapist's technique in a typical meeting is similar to that generally used at the clinic. He stimulates interaction, reflects expressed feelings, and compares and summarizes the various experiences, attitudes, and feelings expressed by different members. Where possible, he connects related material, including cause and symptomatic effect. Although he is generally inactive during the group discussion, he demonstrates continual interest and acceptance.

However, in several ways the therapist's technique differs from that used in closed groups. The therapist is more active in introducing new members, and in describing (repeatedly because the group membership changes) what group psychotherapy is about. In most closed groups the therapist purposely avoids such orienting remarks. Because the patients are not expected to remain in the group for very long, focus is mostly on what sustains, impairs, and improves individual adaptation, and the relationship to symptoms. There is less activity and depth in interpretation than would be the case in a closed group.

TABLE 1

PERCENTAGE OF PATIENTS IN E AND C GROUPS REMAINING IN PSYCHOTHERAPY UNTIL CUTOFF DATE\*

Group	Sessions						
	1	5	10	20	30	40	50
E	100	81	62	49	26	17	9
C	100	66	46	27	15	6	4

\* The number of sessions a patient could theoretically attend until cutoff date varied depending on when he began therapy during the course of the study. Thus the low percentages indicated for 30, 40, or 50 sessions is partly due to the small number of cases in each group who had entered early enough to attend that many sessions by cutoff date. These percentages are not normative data on therapy stay-rate.



One other difference lies in the therapist's willingness to see any patients for brief individual contacts before or after group meetings if he deems it advisable. Patients seek out the therapist, perhaps to tell him of some recent disturbing event in their lives. The opportunity is then taken to accept this as a problem and encourage the patient to bring the problem to the group for discussion, emphasizing that this is exactly the purpose of the group. The therapist, too, will ask one or more patients to stay after a meeting to discuss with them individually how they see treatment progressing, and to tell them some of his observations of their behavior. In this way, silent members in particular are recognized and their difficulties in participating are discussed with them. Often exploration of their feelings and encouragement to participate has resulted in lessened guilt, stronger group feeling, and increased participation in meetings.

Aside from the special effects of the therapist's activity, the effects of two other factors should be considered. For one thing, the open-ended group meetings were presented to the patient as a trial situation. No immediate commitment to psychotherapy was promised or demanded, whereas this is often the case in direct referral to a therapist. A new patient applies for treatment at the clinic at a time when his anxiety and self-concern are often at their highest points, defenses are fluid, and he may have mobilized himself to do something constructive about it. This is the best time to initiate treatment. At the same time the patient (at our clinic, at least) is a person of limited acquaintance with psychotherapeutic practice, is not psychologically minded, and sees no special merit in any kind of extended psychological treatment. He wants help but at this time could only want an impermanent arrangement. That is exactly what is offered by the open-ended group. Also the prospect is offered that if the trial treatment is successful, he and the therapist together can make the most suitable arrangements for more permanent treatment.

This atmosphere of no immediate commitment can have a beneficial effect on the therapist as well as on the patient. It relieves

him of the subtle pressure to retain a case in treatment once it has been screened by Intake and assigned to him as appropriate for treatment. The therapist's acceptance of the appropriateness of psychotherapy for his new patient often exceeds the patient's acceptance of this procedure which is alien to him.

A feature unique to the open-ended group which undoubtedly added to its effectiveness was its method of transferring patients in pairs or other groupings to ongoing or newly formed groups. These men were probably helped in adjusting to transfer by the strengthening presence of their old group associates, although the groups assuming these patients then had problems involving assimilation of clique subgroups.

As a result of the special conditions of high group turnover and minimum screening of new patients for the open-ended group, we were able to make some observations of the effects of such turnover and of the problems resulting from heterogeneous group membership.

Only one element of variability among the group members was found to result in a clear-cut obstacle to group progress. That was including in the group at the same time psychotics with neurotics and patients with character disorders. This difficulty had been anticipated and Intake personnel were instructed to exclude the psychotic patient from referral to this group. However, some were mistakenly admitted. Invariably this led to severe difficulty. The psychotic members' distortions and projections served to stampede neurotic patients with weak defenses into distorted thinking and made the others distinctly uneasy and silent.

While other elements of variability among group members, such as age, education, occupational level, ethnic background, history of criminalism, or previous experience in psychotherapy, created problems or issues, these served as useful group-discussion material and did not seriously hamper the group work.

As for the frequent turnover in group membership, this special group, with its expressed policy of turnover, did not develop the same intensity of angry feelings when

new patients were admitted as do closed groups. The individual member's role in this group changed rapidly from that of newcomer to old member because, relatively soon after he entered the group, the older, perhaps more active members, departed and more new members entered. When members left, the effect seemed to be a closing of ranks. However, departure of patients was also seen as a "graduation" and left some remaining patients feeling deprived or unrecognized.

How were patients selected from the group for discharge or for transfer to closed groups or individual treatment? Most of those discharged were self-terminated; they just stopped coming to meetings. Three, as already mentioned, were deemed not to need further treatment. Readiness for transfer of a group member to further treatment was determined by his degree of acceptance of psychotherapy as evidenced in his attendance, degree of participation in discussion, and expressed opinions about treatment.

Transfer to individual treatment seemed indicated for those patients who had acute problems obviously tied to recent events in their lives or to other environmental pressures. It also seemed indicated for those patients whose homosexual anxiety became overpowering in a group or for those who were so narcissistic as to be unable to relate in a group even after as much as 6 months trial.

Transfer to closed groups was indicated for those patients with chronic difficulties in social adjustment and for those who felt stronger and more secure in a group setting than in individual therapy, for in a group they could share the therapeutic work.

A very important element in the constant selection and transferring of patients was the willingness of the therapist to be constantly "losing" his best patients, retaining only those whom he had not yet been able to properly evaluate or prepare for further treatment.

#### A REPLICATION

The effects of two important factors on the procedure still remained untested. These

were (a) the use of only one therapist for the experimental group and (b) a greater proportion of patients in the experimental group going on to closed group therapy while the control group had a greater proportion in individual therapy. We have replicated the study with another therapist and have investigated the average length of stay in closed group therapy in our clinic as opposed to individual therapy. Although Nash, Frank, Gliedman, Imber, and Stone (1957), report a much higher dropout rate in group therapy than individual therapy, we checked for this at our clinic and found no difference. During an 8-month period, we sampled 36 patients who began individual therapy and compared their attendance over a 12-month period with 10 patients assigned to group therapy. There was no difference in dropout rate.

We repeated the experiment of the open-ended group for an 8-month period, using another therapist. During this period, of 31 patients who were referred to the open-ended group, 22 attended for at least one meeting. In a manner similar to the earlier study, we selected 25 control patients who appeared for at least one therapy session. The experimental patients stayed an average of 7.7 meetings in the open-ended group and then went on to 10.2 meetings in further therapy. Thus, they were in treatment 17.9 sessions. This is counted up to a cutoff date 13 months after the start of the open-ended group. In comparison, the 25 control patients stayed for a total of 15.8 sessions. This minimal difference is deceptive. The first batch of seven patients in the open-ended group, the therapist transferred out very quickly (they averaged only 3.5 meetings in his group). These patients did not stay on in treatment very long. Their total treatment sessions were only 7.5. Realizing this, the therapist kept his later patients in open-ended group longer, and the next consecutive 15 patients averaged almost 10 meetings in his group and then stayed on for a total of almost 23 sessions by the cutoff date. Thus, if we compare the average of 23 to the 15.8 sessions of the control group, we have a similar favorable result, as was the case with the original study. We found that patients



should be kept 10 sessions rather than just a few for best effect.

## REFERENCES

- BITTERMAN, C. M. Serving applicants when there is a waiting list. *Soc. Casewk.*, 1958, 39, 356-360.
- GINOTT, H. G. Group screening of parents in a child guidance setting. *Int. J. group Psychother.*, 1956, 6, 405-409.
- GINOTT, H. G., & BLEK, L. A study in non-attendance of initial interviews in a community clinic. *Int. J. group Psychother.*, 1959, 9, 314-321.
- HOTKINS, A. S., KRIEGSFELD, M., & SANDS, R. An interview group therapy program for the waiting-list problem. *Soc. Wk.*, 1958, 3, 29-34.
- KURLAND, S. H. Length of treatment in a mental hygiene clinic. *Psychiat. Quart. Suppl.*, 1956, 30, 83-90.
- NASH, E. H., FRANK, J. D., GLIEDMAN, L. H., IMBER, S. D., & STONE, A. R. Some factors related to patients remaining in group psychotherapy. *Int. J. group Psychother.*, 1957, 7, 264-274.
- PECK, H. B. An application of group therapy to the intake process. *Amer. J. Orthopsychiat.*, 1953, 23, 338-349.
- STONE, A. R., PARLOFF, M. B., & FRANK, J. D. The use of "diagnostic" groups in a group therapy program. *Int. J. group Psychother.*, 1954, 4, 274-284.
- TABOROFF, L. H., BROWN, W. H., KORMER, I. N., REISER, D. E., TALMADGE, M., GOATES, B. L., & STEIN, E. A note on intake diagnosis in groups. *Int. J. group Psychother.*, 1956, 6, 193-196.

(Received June 18, 1962)

## NOTES AND COMMENTS

### COST EFFICIENCY AND TEST PREDICTION

DAVID RIMM

Stanford University

In evaluating the efficiency of a given screening instrument it is essential that the cost of a false positive error, relative to the cost of a false negative error, be taken into consideration whenever this information is available. An efficiency index based solely upon the expected proportion of correct classifications might well lead to an incorrect decision regarding the acceptance or rejection of a given test. A relatively simple formula is presented, taking into account error costs, as well as the discriminatory power of the instrument, and the population base rate. The formula is then applied to certain hypothetical examples, showing how a decision may be reversed when information regarding error costs is introduced.

In an informative and thought-provoking article, Meehl and Rosen (1955) considered the effectiveness of the more traditional assessment technique (i.e., standardized test) in comparison with a selection procedure involving the assignment of all subjects to the same class or category. One of the main points of their paper was that quite frequently even the better selection instruments are inferior, in terms of total correct decisions, to the procedure of simply passing everyone. The purpose of the present paper is to specify further the value or efficiency of a given selection test by taking into account the cost of the two possible types of errors. Thus, false positives may be considerably more costly, in terms of dollars, or man-hours, than false negatives, or vice versa.

As an illustration, suppose a psychiatric test is set up in order to screen out all potential neuropsychiatric "washouts" in a flight training program. Suppose that, on the average, it costs \$10,000 to put a man through flight training. If, at the close of such training a man presents signs of a neuropsychiatric breakdown, he will immediately be discharged. The price for such a selection error (a false negative) is \$10,000 (less if the breakdown occurs prior to the conclusion of training). On the other hand, suppose that the cost of recruitment and induction is approximately \$1,000. An individual who is rejected at this stage, but who is actually quite resistant to the types of stress one experiences in flight training (a false positive), represents the second type of selection error, the price of this mistake being \$1,000. The ratio of the cost of these two types of errors is 1:10. In other words, it is much better to reject a fit individual than to

accept or pass a man who is, in reality, unfit for such training.

Consistent with Meehl and Rosen, the following notation will be used throughout this presentation:

$p_1$  is the proportion of sick people called "sick" by the test (i.e., the proportion of *true positives*).

$p_2$  is the proportion of healthy people incorrectly labeled "sick" by the test (i.e., the proportion of *false positives*).

$P$  is the actual proportion of sick people in a given population (i.e., the *base rate*).

For  $N$  individuals,  $NP$  of them will actually be sick, and  $NP(1 - p_1)$  will incorrectly be labeled "well" by the test. Thus there will be a total of  $NP(1 - p_1)$  false negatives. For the same  $N$  individuals,  $N(1 - P)$  will actually be healthy and  $N(1 - P)p_2$  will incorrectly be labeled "sick." There will thus be a total of  $N(1 - P)p_2$  false positives.

Let " $A$ " be the cost of each false negative, and " $B$ " equal the cost of each false positive error. The total cost, for both types of errors for any sample of size  $N$  is then given by:

$$\text{Total Cost} = ANP(1 - p_1) + BN(1 - P)p_2.$$

If, instead of using a selection test, *all* individuals are passed or called "healthy," there will be a total of  $NP$  selection errors (all false negatives), and the cost for these errors will be given by:

$$\text{Base Cost} = APN.$$

We can now define "Cost Efficiency" as the proportion decrease in cost when a selection test



is introduced. Thus:

$$\text{Cost Efficiency} = \frac{\text{Base Cost} - \text{Total Cost}}{\text{Base Cost}}.$$

$$\text{Cost Efficiency} = p_1 - \frac{B(1 - P)p_2}{AP}.$$

If the ratio of the cost of a false positive to the cost of a false negative is set equal to  $R$ , then:

$$\text{Cost Efficiency} = p_1 - \frac{R(1 - P)p_2}{P}.$$

A few examples should aid in the interpretation of the Cost Efficiency. In the hypothetical flight training example " $R$ " was equal to  $\frac{1}{10}$  or .1. Suppose that on the average 25% of the men "wash out" during the critical last week of training. Suppose that a screening procedure is now introduced which enables one to identify correctly 60% of the potential washouts. But at the same time this procedure incorrectly labels 20% of the potentially qualified men as unfit for training. From the above formula, the Cost Efficiency for this selection test would be equal to:

$$.6 - \frac{.1(1 - .25).2}{.25} = .54.$$

This means that for every dollar that would have been spent paying for errors resulting from the procedure of accepting each applicant, .54 dollars will be *saved* as a result of using the screening test. To put it another way, for each dollar spent as a result of errors associated with passing everyone, only  $1 - .54 = .46$  dollars will be *spent* if the screening test is used instead.

There may be situations where the cost of training is low in comparison with the cost of recruitment. This might be the case when a company is confronted with the task of hiring semi-skilled construction workers for an extremely undesirable job overseas. Here individuals might be selected on the basis of whether or not they would remain on the job for a certain period of time, since every applicant is assumed to possess enough skill to handle the job. Suppose that on the average 50% of the men quit the job before they have completed 2 years work. Now suppose that a selection procedure is introduced which is capable of correctly identifying 70% of the "quitters" but mislabels 25% of the "nonquitters." Finally, suppose that in the long run, rejecting a nonquitter will be four times as expensive as accepting a quitter.

The Cost Efficiency for this selection test would be equal to:

$$.7 - \frac{4(1 - .5).25}{.5} = -.30.$$

A negative Cost Efficiency has the same interpretation as a positive index. In this instance, for each dollar spent due to errors resulting from hiring every man,  $1 - (-.30)$  or 1.30 dollars will be spent if the screening procedure is used instead. In this example it would be much more economical simply to accept every applicant.

If " $R$ " is set equal to unity, then the comparison which the Cost Efficiency measure makes would be reduced, in effect, to the sort of comparison Meehl and Rosen considered in their paper. In our flight training example an  $R$  of 1 would have resulted in a Cost Efficiency of 0.00. Since a zero Cost Efficiency implies no differential advantage, the test would have been rejected, because administering any instrument costs a certain amount of money. In the second example, an " $R$ " of unity would have resulted in a Cost Efficiency of .45, so that the test would probably have been accepted as part of the standard induction procedure. Note that in both examples, ignoring relative error costs leads to a decision which is *opposite* to that which would have been made had cost been considered.

As has been indicated, whenever the Cost Efficiency for a given instrument is positive, the instrument has the advantage over base rate prediction. The maximum value of the Cost Efficiency is unity, possible only when the proportion of true positives is equal to 1 and when either the base rate is equal to 1 or there are no false positives. The minimum value for the index is negative infinity, with the index tending toward this value as the base rate approaches zero. Since the slope of the curve relating Cost Efficiency and base rate is always positive (see Fig. 1), an

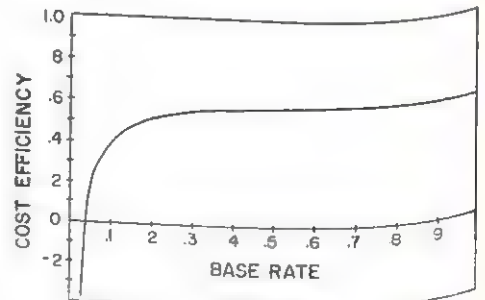


FIG. 1. Cost Efficiency as a function of base rate,  $P$ . (The slope is given by  $Rp_2/P^2$ ; the  $P$  intercept equals  $Rp_2/p_1 - Rp_2$ . This particular curve is for the flight training example.)

increase in base rate will necessarily mean an increase in Cost Efficiency, all other things being equal.

### DISCUSSION

Through this development we have made certain implicit assumptions. One such assumption is that the cost of administering a screening test is negligible. In most instances this would probably be true. Another assumption is that all individuals who pass the test will be accepted. In many cases a selection ratio is involved such that only a certain fixed number of openings is available. Since the Cost Efficiency would tend to vary with the number of applicants, this index would have to be adjusted so as to take this into consideration. This might be accomplished by expressing both  $p_1$  and  $p_2$  as functions of the cutting score.

Whenever these assumptions can be met, the Cost Efficiency can be calculated, providing of course that knowledge of the base rate and error

costs is available. Unfortunately, as Meehl and Rosen have pointed out, this sort of information is frequently nonexistent. In many situations, the cost of errors could never be measured in dollars and cents (e.g., an unfit pilot dying in a plane crash, or a mental patient being refused therapy, though in fact he is quite ill and experiencing considerable anguish). However, some sort of relative valuation of the types of errors will almost always be possible. If this judgment can somehow be translated into a number ( $R$ ), no matter how cold-blooded this procedure may seem, it would most certainly be better than assuming that all types of mistakes are equally bad.

### REFERENCE

- MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, 52, 194-216.

(Received May 15, 1962)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 1, 91-94

## ON THE DETECTION OF WILLFUL FALSIFICATIONS IN THE MMPI

JOHN E. EXNER, JR.<sup>1</sup>

*Bowling Green State University*

EUGENE McDOWELL, JOAN PABST, WILLIAM STACKMAN, AND LYNN KIRK

*DePauw University*

It seems quite evident that the patterning of the MMPI profile can be significantly altered by the willful intent of the subject since it is a self-report device. Ellis (1953) cites 22 of 25 studies which demonstrate the device to be fable. Grayson and Olinger (1957) found that psychiatric patients asked to respond to the MMPI as would a "normal" individual were able to produce greatly improved profiles when compared to original MMPI results obtained from each patient upon admission to the hospital. Thus the total usefulness of the device as a diagnostic instrument, particularly when used for screening or in situations where social desirability is an important element, seems to be dependent largely upon the honesty of the subject or on the ability of the interpreter to detect willful attempts at dishonest reporting.

The construction of the MMPI includes four

"validity scales" in order to provide the interpreter with some basis for the detection of fraudulent profiles. There has been some support for the usefulness of these validity scales. Cofer, Chance, and Judson (1949) found that college students attempting to malingering will have significantly high  $F$  scores while those attempting to fake normal can be detected by an additive combination of the  $L$  and  $K$  scores. Calvin and McConnell (1953) refer to eight studies in which faking on the MMPI can be detected at a statistically significant level using the validity scales. Recently, however, other studies have suggested that a high score on a single validity scale does not necessarily invalidate the diagnostic patterning. Gynther (1961) has demonstrated that a high  $F$  score is found in almost all cases of behavioral disorders. Heilbrun (1961) reports evidence suggesting that a high  $K$  score can only be interpreted in the light of the psychological adjustment of the individual. The use of a linear

<sup>1</sup> Formerly Director, Bureau of Testing and Research at DePauw University.



combination of scores to detect malingering on the MMPI was first suggested by Gough (1947, 1950) who reports that the formula  $F$  minus  $K$  satisfactorily identifies a substantial portion of malingered profiles when a cutoff score of  $+9$  is used. Hunt (1948) using Gough's " $F-K$  Dissimulation Index" reports similar success using a cutoff score of  $+10$ . Both Gough and Hunt attempted to find a cutoff score for the  $F-K$  index which would satisfactorily detect "faking good" profiles but report little success as considerable overlap exists between  $F-K$  score distributions derived from faking good profiles and "honestly reported" profiles. Gough (1954) has also found that there are 74 items in the MMPI to which persons attempting to malingering respond significantly different than do diagnosed psychoneurotics. Gough refers to this grouping of items as a "Dissimulation Scale" ( $Ds$ ).

The purpose of the present study is to investigate methods, including the  $F-K$  index,  $Ds$  scale,  $L+K$  index, and a variety of unexplored linear combinations of scores, that might be useful in distinguishing between malingered, faking good, and honestly reported profiles obtained in a screening situation.

#### PROCEDURE

Fifty college students, including 25 males and 25 females ranging in age from 20 to 22 years, were used

as subjects. They were divided into two groups on the basis of similarity of academic grade-point average. One group contained 13 males and 12 females, while the second was comprised of 12 males and 13 females. In the first part of the investigation the subjects in one group, Group G, were asked to respond to the MMPI in such a manner as to appear normal or socially desirable as would an attractive job or school applicant. The subjects in the second group, Group M, were asked to respond to the MMPI in a manner as to appear sufficiently deviant to be exempt from some social responsibility such as military service but not so deviant that institutionalization would be required. In the second portion of the investigation all 50 subjects were asked to respond to the MMPI again but this time in a completely honest manner as if they were interested in gaining information concerning themselves. Whereas names were requested in the first administration of the MMPI when the subjects were attempting to falsify profiles, complete anonymity was assured in the second administration and only data concerning sex and group classification were recorded in an effort to insure the honesty of reporting.

#### RESULTS

All profiles were scored for all nine diagnostic scales and the four validity scales plus the  $F-K$  index. All records from Group M were also scored on the  $Ds$  scale and the  $L+K$  index was computed for all records from Group G. Calcula-

TABLE 1  
COMPARISONS OF MMPI MEAN STANDARD SCORES,  $Ds$  SCORES, AND SPECIAL LINEAR COMBINATIONS OF RAW SCORES FOR THE HONESTLY REPORTED AND WILLFULLY ALTERED RECORDS OF EACH GROUP

Scale	Group G ( $N = 25$ )					Group M ( $N = 25$ )				
	Honest		Faked		$t$	Honest		Faked		$t$
	$\bar{X}$	$SD$	$\bar{X}$	$SD$		$\bar{X}$	$SD$	$\bar{X}$	$SD$	
$L$	43.32	5.77	50.64	8.16	3.66*	43.72	4.15	47.16	10.14	1.58
$F$	51.68	7.26	47.72	5.08	2.24**	55.28	6.82	100.80	17.65	12.07*
$K$	58.20	7.42	62.92	6.33	2.42**	57.44	7.67	43.64	5.04	7.58*
$Hs$	48.31	6.74	48.44	7.14	0.07	47.50	7.62	70.44	18.86	5.65*
$D$	48.77	9.90	44.72	5.12	1.84	53.15	6.75	81.96	19.40	7.03*
$Hy$	55.38	7.42	56.20	4.12	0.49	53.19	6.27	67.04	15.26	4.21*
$Pd$	54.96	7.42	51.04	4.84	2.23**	57.50	7.60	85.32	13.12	9.24*
$Pa$	53.58	8.05	53.32	6.61	0.13	51.38	9.23	79.28	18.79	6.69*
$Pl$	53.65	8.60	50.80	4.74	1.48	54.96	7.42	85.32	13.12	10.12*
$Sc$	55.65	7.49	54.20	4.47	0.84	57.04	8.64	100.00	19.47	10.11*
$Ma$	55.77	8.37	59.36	13.55	1.13	59.77	12.19	75.44	14.52	4.17*
$Ds$	—	—	—	—	—	10.62	5.57	40.27	11.76	11.63*
$F-K$	-13.28	5.68	-17.44	5.20	2.70*	-11.00	6.23	22.25	11.41	12.84*
$L+K$	18.80	4.89	23.40	4.96	3.31*	—	—	—	—	—

\*  $p < .01$ .

\*\*  $p < .05$ .

tion of  $t$  tests was made comparing the willfully-altered and honestly-reported records for all scales and special score combinations. Some of the more pertinent statistical data is shown in Table 1. From examination of Table 1 it will be noted that a statistically significant difference at .01 exists for all of the regular MMPI scales, except the  $L$  scale, when the Group M honestly-reported and malingered records are compared indicating that the attempt at willful alteration was successful. It should also be noted that considerably fewer significant differences exist between the honestly-reported and faking-good records of Group G. In this instance a difference at .01 exists only for the  $L$  scale and differences at .05 exist for the  $F$ ,  $K$ , and  $Pd$  scales suggesting considerable similarity between the records taken from Group G.  $t$  tests comparing each set of willfully-altered records with all 50 honestly-reported records yield statistically significant differences essentially the same as those found when the willfully-altered and honestly-reported profiles of a single group were compared.

It should be noted in Table 1 that a difference at .01 exists between the honestly-reported and malingered records of Group M on the  $Ds$  scale. A frequency distribution of these scores reveals a range for honestly-reported records of 0 to 26 with 24 of the 25 scores falling below 20. In contrast, the range for malingered records is 20 to 64. The  $F-K$  index also shows a difference at .01 in Group M. Using Hunt's cutoff of +10, 17 of the 25 malingered records can be detected whereas a cutoff of +12 detects 24 of the 25. The  $F$  scale taken alone is most effective in detecting malingered records. No malingered record yielded an  $F$  score of less than 12, the range being from 12 to 30 while none of the 50 honestly-reported records yielded an  $F$  score of more than 11, the range being 0 to 11. The detection of faking-good records was considerably less successful. Although a statistically significant difference of .05 for the  $F$  and  $K$  scales and of .01 for the  $L$  scale and  $F-K$  and  $L+K$  indexes exists between the honestly-reported and faking-good records of Group G, considerable overlap in the ranges of scores which eliminates the usefulness of these items in terms of practical application. The honestly-reported records of Group G yield an  $F-K$  range of +5 to -22 while the faking-good records have a range of -3 to -23. The Group G distributions for the  $K$  scale show a range of 9 to 23 for honestly-reported records and 9 to 24 for faking-good records and for the  $F$  scale, 0 to 14 and 0 to 12, respectively. The  $L$  scale distributions for Group G are 0 to 6 for honestly-reported records

and 0 to 9 for faking-good records with considerable overlap. Likewise, the distributions of the  $L+K$  scores show too much overlap to be practically useful in detecting faking-good records with the honestly-reported distribution ranging from 12 to 29 and a faking-good range of 11 to 30. Not shown in Table 1 are a variety of linear combinations which were evaluated with regard to the detection of faking-good records. These include  $2L+K$ ,  $F-2L$ ,  $K+D$ ,  $K+Pd$ , and  $2L+Pd$ . None of these combinations produced satisfactory frequency distributions whereby more than 65% of the faking-good records could be accurately discriminated from the honestly reported records although all revealed statistical significance at least at .05.

### DISCUSSION

The results of this investigation lend support to the usefulness of Gough's  $Ds$  scale and  $F-K$  dissimulation index as methods of detecting malingering. This data also supports Cofer's (Cofer et al., 1949) findings that the  $F$  scale taken alone can be useful for the detection of malingered records. In contrast, this data does not support Cofer's findings that a  $L+K$  index will satisfactorily detect faking-good records nor the more general proposition that the  $K$  scale taken alone is practical for this purpose. On the contrary, the results suggest that deliberate attempts at faking "normal" profiles can be successful and are reasonably undetectable, at least through the use of any single scale or linear combination of raw scores. The question of whether the honestly-reported profiles truly represent sincere reporting must, of course, be considered. All evidence seems to indicate that they do. For instance, the honestly-reported records of Group G are highly similar to those given by Group M even though the previous records of these groups were completely different. In addition, significant changes did take place in the scores of the  $L$ ,  $F$ , and  $K$  scales of the honestly-reported records when compared with the faking-good profiles. Finally, great pains were taken to insure anonymity to the subjects when the honestly-reported records were requested, and the intent of the study was discussed with the subjects prior to the first MMPI administration so that there appears to be no realistic reason to shed doubt on the validity of their responses. Assuming that the data collected are valid, it can only be concluded that the usefulness of the MMPI as a screening device seems seriously limited by the fact that willful manipulation of the record can be easily accomplished. When such manipulation is in the



direction of creating a "socially desirable" profile there appears to be no way of detecting this manipulation with reasonable confidence. Of course, this study has not considered the element of total configurational signs as evidence of willful alteration. However, the number of actual linear combinations attempted was such as to include many of the diagnostic scales although to no practical avail. Thus it would seem reasonable to suggest that conclusions concerning willful alternation in the direction of creating socially desirable profiles based on configural signs be made with definite caution. It also seems worthwhile to note that the *K* scores of the honestly-reported records used in this study tend to be relatively high, which in turn obviously suppress the interpretive value of the *K* scale as a general measure of defensiveness. Since the subjects in this investigation are college students oriented toward intellectual sophistication it might be hypothesized that the high mean *K* scores represent a set toward social desirability which might not evidence itself so intensely in a population manifesting a greater range of intelligence.

## REFERENCES

- CALVIN, A., & McCONNELL, J. E. Ellis on personality inventory. *J. consult. Psychol.*, 1953, 17, 462-464.
- COFER, C. N., CHANCE, JUNE, & JUDSON, A. J. A study of malingering on the Minnesota Multiphasic Personality Inventory. *J. Psychol.*, 1949, 27, 491-499.
- ELLIS, A. Recent research with personality questionnaires. *J. consult. Psychol.*, 1953, 17, 45-49.
- GOUGH, H. G. Simulated patterns on the MMPI. *J. abnorm. soc. Psychol.*, 1947, 42, 215-225.
- GOUGH, H. G. The *F* minus *K* Dissimulation Index for the MMPI. *J. consult. Psychol.*, 1950, 14, 408.
- GOUGH, H. G. Some common misconceptions about neuroticism. *J. consult. Psychol.*, 1954, 18, 287-292.
- GRAYSON, H. M., & OLINGER, L. B. Simulation of normalcy by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, 21, 73-77.
- GYNTHIER, M. D. The clinical utility of invalid MMPI *F* scores. *J. consult. Psychol.*, 1961, 25, 540-542.
- HEILBRUN, A. B. The psychological significance of the MMPI *K* scale in a normal population. *J. consult. Psychol.*, 1961, 25, 486-491.
- HUNT, H. F. The effect of deliberate deception on MMPI performance. *J. consult. Psychol.*, 1948, 12, 396-402.

(Received June 18, 1962)

# JOURNAL OF CONSULTING PSYCHOLOGY

April 1963

Vol. 27, No. 2

## Contents

### ARTICLES

Need for Approval and the Premature Termination of Psychotherapy: <i>Bonnie R. Strickland and Douglas P. Crowne</i>	95
Intelligence Test Performances of Brain Damaged Subjects with Lateralized Motor Deficits: <i>Homer B. C. Reed, Jr. and R. M. Reitan</i>	102
Social Desirability under Role Playing Instructions: A Reply to Walker: <i>Jerry S. Wiggins</i>	107
MMPI Differences Associated with Sex, Race, and Class in Two Adolescent Samples: <i>Robert L. McDonald and Malcolm D. Gynther</i>	112
Interviewer Behavior as a Function of Standardized Client Roles: <i>Kenneth Heller, Roger A. Myers, and Linda Vikan Kline</i>	117
Language Predictability as a Function of Psychotherapeutic Interaction: <i>Stanley Feldstein and Joseph Jaffe</i>	123
Relative Accuracy of Actuarial Prediction, Experienced Clinicians, and Graduate Students in a Clinical Judgment Task: <i>Lawrence C. Grebstein</i>	127
Purdue Pegboard as a Predictor of the Presence and Laterality of Cerebral Lesions: <i>Louis D. Costa, Herbert G. Vaughan, Jr., Eric Levita, and Norman Farber</i>	133
Empathy, Need to Change and Improvement with Psychotherapy: <i>Rosalind Dymond Cartwright and Barbara Lerner</i>	138
Intraindividual Consistency in "Creative" and "Memory" Stories Written for TAT Pictures: <i>Ruth C. Wylie, Boyd D. Sisson, and Earl Taulbee</i>	145
Evidence Regarding the Equivalence of Ipsative and Normative Personality Scales: <i>Alfred B. Heilbrun, Jr.</i>	152
Hypnotic Susceptibility and MMPI Profiles: <i>Robert E. Schulman and Perry London</i>	157
Revision of the MMPI K Correction Procedure for Improved Detection of Maladjustment in a Normal College Population: <i>Alfred B. Heilbrun, Jr.</i>	161
Relative Importance of Test Item Content: <i>Warren T. Norman</i>	166
Client Likability: A Variable in the Study of Psychotherapy: <i>Norton Stoler</i>	175

### NOTES AND COMMENTS

Psychoticism or Ten Psychotic Syndromes? <i>H. J. Eysenck</i>	179
Canonical Variates and Second-Order Factors: A Reply: <i>Maurice Lorr, Douglas M. McNair, C. James Klett, and Julian J. Lasky</i>	180

### BRIEF REPORTS

Intraindividual Verbal-Numerical Discrepancies and Personality: <i>Arnold E. Dahlke and Richard H. Dana</i>	182
Discrimination within the Delinquency Continuum on Gough's Socialization Scale: <i>Gaylord L. Thorne</i>	183
Type of Brain Damage and Intellectual Functioning in Children: <i>Daniel V. Caputo, William E. Edmonston, Jr., Luciano L'Abate, and Samuel R. Rindberg</i>	184
Judgments of Motivation for Psychotherapy: Some further Explorations: <i>Martin S. Wallach</i>	185
A Further Follow-Up Comparison of Psychotherapeutic Programs: <i>George W. Fairweather and Ralph Simon</i>	186



# Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specification given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

## Specifications

**Brief Report.** The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the

author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

<sup>1</sup> An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Documentation No. — from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$— for microfilm or \$— for photocopies, make checks payable to: Chief, Photoduplication Service, Library of Congress.

**Extended report.** Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (a) The extended report should be typed with single spacing for economy in duplication. (b) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (c) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.

## NEED FOR APPROVAL AND THE PREMATURE TERMINATION OF PSYCHOTHERAPY<sup>1</sup>

BONNIE R. STRICKLAND

AND

DOUGLAS P. CROWNE<sup>2</sup>

*Emory University*

*Ohio State University*

This study tested the hypotheses that approval-dependent individuals (a) tend prematurely to terminate psychotherapy, and (b) are rated by therapists as more defensive and less improved than patients less approval-dependent. 85 psychiatric outpatients completed the Marlowe and Crowne Social Desirability scale (the measure of need for approval) and rated their improvement in psychotherapy. Therapists rated 30 of these patients on defensiveness, attitude towards patient, patient's attitude towards therapist, improvement, and satisfaction with therapy. Additional measures included diagnosis, social class, and ordinal position. Results confirmed the hypotheses. The high need for approval group terminated significantly ( $p < .005$ ) earlier. Approval-motivated patients were generally given more negative ratings by therapists. Approval-motivated females rated themselves as more improved. Ordinal position and social class failed to predict stay in therapy. The implications of these findings were discussed.

In a series of recommendations on the conduct of psychoanalysis, Freud (1949) observed that the premature termination of treatment could be attributed to the initial unsuitability of the patient for psychoanalysis (e.g., inadequate intelligence, low social class, psychosis) or to an early and dramatic manifestation of resistance. He advocated a period of trial analysis in order to permit the analyst to assess the patient's motivation for therapy and whether or not he met the (then extant) criteria for treatment.

Contemporary research on early termination has, in the main, taken a straight empirical approach to the problem. A not inconsiderable body of studies in the last decade has established, with certain inconsistencies and contradictions, two major predictors of abrupt termination. The first class of these variables roughly corresponds to Freud's criterion of unsuitability: education (Sullivan, Miller, & Smelser, 1958), social class (Auld & Myers, 1954; Hollings-

head & Redlich, 1958; Rosenthal & Frank, 1958), and intelligence (Affleck & Mednick, 1959; Auld & Eron, 1953). The second set of predictors is probably best conceptualized under the heading of defensiveness. Indices such as the Rorschach, MMPI, semiprojective sentence completions, and therapists' ratings have shown differences between patients who prematurely terminate and those who continue in psychotherapy. Unproductive Rorschach records (Affleck & Mednick, 1959; Gallagher, 1953; Gibby, Stotsky, Hiler, & Miller, 1954; Taulbee, 1958), more limited acknowledgment of symptoms and of personal dissatisfaction on the MMPI (Taulbee, 1958), sentence completions indicative of evasion and unwillingness to reveal oneself (Hiler, 1959), and a more limited prognosis as rated by therapists (Garfield & Affleck, 1961) appear to be characteristic of the abrupt terminator. These findings are further consonant with the essential meaning of Freud's early clinical observations.

The present study was undertaken to test a theoretically derived prediction concerning the phenomenon of defensive early termination of psychotherapy. The major hypothesis was that patients characterized by a high need for approval are more likely to terminate psychotherapy early. As a consequence of their need to maintain and defend a vulner-

<sup>1</sup> Portions of this study were presented at the Midwestern Psychological Association meetings, Chicago, May 1962.

<sup>2</sup> We would like to express our gratitude to the therapists and patients who graciously and uncomplainingly completed the various measures required of them. Our thanks are also extended to Shephard Liverant, David Marlowe, and Julian Rotter, whose suggestions contributed both to the analysis and interpretation of the data of this study.



able self-image, these individuals are more likely to avoid the threats associated with personal disclosure (and, possibly, anticipated social censure for seeking this form of help with personal problems) by breaking off therapeutic contact. Previous research on the need for approval has found approval-motivated individuals to be more compliant, persuasible, eager to please, and conforming in a variety of situations: opinion conformity (Marlowe & Crowne, 1961), social conformity (Crowne & Liverant, 1963; Strickland & Crowne, 1962), verbal conditioning (Crowne & Strickland, 1961; Marlowe, 1962), perceptual defense (Barthel & Crowne, 1962), and attitude change (Salman & Crowne, 1962). The present hypothesis is based on an extension of the meanings of the construct of need for approval from an earlier conceptualization emphasizing compliance and influencibility (cf. Crowne & Strickland, 1961). Recent work by Conn (1962) and Crowne and Liverant (1963) suggests that individuals with a high need for approval are more defensive and are concerned with avoiding threats to self-esteem. This proposition receives further support from the initial finding of greater test taking defensiveness by persons with high scores on the Marlowe-Crowne Social Desirability (*M-C SD*) scale, the index of the need for approval.

#### METHOD

The data for the present study were collected as part of a large-scale research project at Columbus Psychiatric Clinic assessing both patient and therapist variables. The therapists were not apprised of the nature of the research, and none of them had any knowledge of the patients' standing on the independent variables.

#### Subjects

Eighty-five patients, in treatment during 1960 and 1961, served as subjects in this study. Throughout the interval spanned by the research, almost every patient seeking help was admitted to psychotherapy which, with only rare exceptions, consisted of weekly therapeutic hours. A majority of the patients were given neurotic or character disorder diagnoses, although a considerable number of them were classified as psychotic or prepsychotic. These diagnoses were made in staff meetings by resident psychiatrists on the basis of intake interviews and, occasionally, psychological test data. The age range of the patient sample was from 20 to 54, with a mean of 33.7 years. The thera-

pists included psychiatrists and psychiatric residents, psychologists, and social workers.

The final sample was comprised of two groups of patients. In Group I were 23 patients, 13 males and 10 females, who terminated psychotherapy during the months of October and November 1960. Group II consisted of 62 patients, 26 males and 36 females. This was a replication sample terminating sometime during May, June, or July 1961.

#### Measures

The index of the need for approval, the major independent variable, was the *M-C SD* scale (Crowne & Marlowe, 1960). The *M-C SD* scale is a 33-item, true-false questionnaire assessing the degree to which individuals avoid self-criticism and depict themselves in improbably favorable terms. An illustrative item is, "I'm always willing to admit it when I make a mistake."

The patients rated their own improvement in therapy on a nine-point scale ranging from improved (1) to unimproved (9). Within the compass of the larger project, therapists were required to rate each of their patients on a slightly modified version of the Seeman (1954) Case Rating scale. Of interest in this study were five items on the Case Rating scale: Number 5, the therapist's estimate of the patient's attitude towards him during therapy; Number 6, the therapist's attitude towards his patient; Number 7, the degree of personal integration versus defensiveness of the patient; Number 9, the therapist's estimate of the patient's satisfaction with the outcome of therapy; and Number 10, the therapist's rating of therapeutic outcome. In contrast to the original, which used nine-point scales, the range of values for the modified version was from 1 to 16.

The closing summaries of treatment in the patients' folders were rated on a five-point scale ranging from optimal improvement and mutual agreement on termination (1) to termination without the therapist's agreement and no improvement (5). Additional variables included education and occupational level, which were combined in a fashion similar to the procedure followed in the Hollingshead Index of Social Position (Hollingshead & Redlich, 1958), and ordinal position. Birth order was included in an attempt to replicate Schachter's (1959) findings that first- as opposed to later-born patients continue longer in psychotherapy. The number of hours of therapy at termination constituted the major dependent variable.

#### Procedure

The *M-C SD* scale was mailed to the patients in Group I within a 3-month interval from the date of termination. In Group II, 33 of the 62 patients were given the *SD* scale at the Clinic. The remaining 29 patients in this group completed the inventory at the conclusion of therapy as in Group I. As shown in Table 1, the *M-C* scale means and standard deviations of the patients in Groups I and II who completed the measure after terminating therapy are closely comparable to the mean and

standard deviation of the Group II patients tested at the Clinic.

The rate of return on questionnaires mailed to the patients in Groups I and II was approximately 70%; in fact, most of the uncompleted questionnaires were undelivered due to incorrect addresses. A Mann-Whitney *U* test revealed no difference in the terminal number of therapy hours of the patients who returned the *SD* scale and those who did not ( $U = 1366.5$ ,  $z = .40$ , *ns*). Thus, it appears unlikely that bias affecting either the independent or dependent variables was introduced into the study by the posttherapy administration of the *SD* scale. Neither does the sample appear to have been biased by selective returns nor were differences observed in *SD* scale scores between patients tested in therapy at the Clinic and patients tested after termination.

Ratings of our improvement in therapy were mailed to the patients within 3 months of termination. Sixty-three of the 85 patients completed their own improvement scale.

The therapists completed the Case Rating scale at varying times, depending on the date each patient began treatment, between November 1960 and May 1961. The therapists' ratings were uniformly made while the patients were in treatment. At the time the ratings were made, the therapists could have had no reliable knowledge of the patients' ultimate dates of termination. Thus, these ratings were not contaminated by the possibility of negative attitudes towards early terminating patients.

## RESULTS

Preliminary to the major analyses, the data were examined for sex and age differences. None, except for the sex difference in patients' ratings of their own improvement reported below, were found, and the analyses were accordingly carried out without regard to sex and age.

To accomplish the test of the central hypothesis, patients' scores on the *M-C* scale were dichotomized at the overall mean (11.87) to yield the high and low need for approval groups. The distribution of the number of hours of therapy of these groups

TABLE 1

MEAN *M-C* *SD* SCORES OF PATIENTS TESTED AT THE CLINIC AND AT THE CONCLUSION OF THERAPY

Group	<i>N</i>	<i>M</i>	<i>SD</i>
I (Posttermination)	23	13.35	6.95
II <sub>a</sub> (At clinic)	34	11.53	6.02
II <sub>b</sub> (Posttermination)	28	11.08	6.76

Note.—*t* (I versus II<sub>a</sub>) = 1.02; *t* (II<sub>a</sub> versus II<sub>b</sub>) = 0.27; *t* (I versus II<sub>b</sub>) = 1.16.

TABLE 2

DISTRIBUTION OF THE NUMBER OF HOURS OF PSYCHOTHERAPY FOR HIGH AND LOW NEED FOR APPROVAL PATIENTS

Hours of therapy	High need for approval ( <i>N</i> = 43)	Low need for approval ( <i>N</i> = 42)
71+ (to 284)	5	11
61-70	1	1
51-60	1	5
41-50	0	5
31-40	7	2
21-30	10	9
11-20	11	5
1-10	8	4

is shown in Table 2. By a Mann-Whitney *U* test, high and low need for approval patients differ significantly in the number of hours of therapy ( $z = 2.84$ ,  $p < .005^3$ ). Looking at the distribution of therapy hours in Table 2, it is clear that above the median hours of therapy, highs are sharply underrepresented, while the reverse is true below the median hours of therapy. Patients low in the need for approval tend to be more equally split above and below the median. At the extremes of the distribution, however, low need for approval patients, in contrast to the highs, are overrepresented in the highest number of therapy hours and underrepresented among the very early terminators. It should be noted that differences of approximately the same magnitude were found in the two samples (Groups I and II); thus, at least within the same clinic, some evidence for the replicability of this finding is afforded.

For the 30 patients in the second group on whom the therapists completed the Seeman scale, correlations were computed between these ratings and the *M-C* scale, the number of hours of therapy, and the index of social class. These correlations, and the intercorrelations of the five Case Rating scale items, are presented in Table 3. Approval-dependent patients tended to receive more negative ratings on each of the five scales. The most striking of these correlations is that between the need for approval scale and therapists' ratings of personal integration versus defen-

<sup>3</sup> Two-tailed test.



TABLE 3

INTERCORRELATIONS OF HOURS OF THERAPY, NEED FOR APPROVAL, AND THERAPISTS' RATINGS

Variable	Variable						
	1	2	3	4	5	6	7
1. Need approval	.68 <sub>a</sub>						
2. Hours of therapy	-.28 <sub>b</sub>						
3. Patient's liking/respect for therapist	-.18	.23					
4. Therapist's liking/respect for patient	-.35	.40*	.66**				
5. Personal integration versus defensiveness of patient	-.67**	.57**	.27*	.38**			
6. Patient's satisfaction with therapy	-.32	.22	.60**	.46**	.55**		
7. Amount of improvement in therapy	-.34	.45*	.46**	.50**	.63**	.79**	
8. Social class	[-.07 N = 30]	.19]	N = 83		.28	.36*	.07]
			N = 30				

<sup>a</sup> Test-retest reliability, 5-6 month interval during which patients were in therapy ( $N = 24$ ).

<sup>b</sup>  $N = 85$ .

\*  $p < .05$ .

\*\*  $p < .01$ .

siveness or disorganization ( $r = -.67$ ). The remainder of the need for approval-therapists' ratings correlations are consistent but fail to reach a conventional level of significance.

Each of the Seeman scales correlated moderately and positively with the number of hours the patient remained in therapy. The index of social class was found to be unrelated to the *M-C* scale ( $r = -.07$ ) and negligibly correlated with length of stay in therapy ( $r = .19$ ). Moderate positive correlations were obtained between social class and therapists' ratings of the patients' liking and respect for the therapist ( $r = .27$ ), satisfaction with therapy ( $r = .36$ ), and the defensiveness scale ( $r = .28$ ). Only the correlation with satisfaction with therapy attains significance ( $p < .05$ ).

Since the patients had been in therapy for varying lengths of time when the therapists' ratings were made (that is, it was not possible to obtain the therapists' ratings after a standard interval in therapy for all patients), it is conceivable that the approval-motivated group was rated lower simply as a function of their tendency to remain less long in therapy. Thus, the negative ratings may simply reflect the fact that their thera-

pists knew them less well. As a test of this possibility, partial correlations were computed between the *M-C* scale and the therapists' ratings with the number of hours the patient had been in therapy at the time of rating partialled out. Table 4 shows these correlations. It is clear that the relationships between the need for approval and the therapists' ratings maintain approximately the same magnitude irrespective of the amount of time the patient had been in therapy at the time he was rated.

Examining the diagnoses of high and low need for approval patients, no differences were found. Psychiatric diagnosis was further unrelated to the other variables of the study.

As a further assessment of progress in, and the outcome of, therapy, the ratings of the closing summaries were compared with certain of the major variables of the study. The interrater reliability of these ratings was .68 ( $N = 25$ ). The terminal ratings correlated .14 with the *M-C* scale,  $-.34$  with hours of therapy,  $-.31$  with the therapists' prior ratings of improvement, and  $-.29$  with therapists' ratings of defensiveness.<sup>4</sup>

<sup>4</sup> The negative sign of the last two correlations is due to the reversed scoring of the scales.

TABLE 4  
PARTIAL CORRELATIONS BETWEEN NEED FOR  
APPROVAL AND THERAPISTS' RATINGS

Therapists' ratings	Partial correlations with need for approval
Patient's liking respect for therapist	-.11
Therapist's liking respect for patient	-.29
Personal integration versus defensiveness of patient	-.61**
Patient's satisfaction with therapy	-.25
Amount of improvement in therapy	-.24

Note.— $N = 30$ .  
\*\*  $p < .01$ .

At termination, 63 patients completed ratings of their own improvement on a scale ranging from 1 (improved) to 9 (unimproved). For males, there was no relationship between ratings of own improvement and the *M-C* scale ( $r = .01$ ). A significant correlation was obtained for females ( $r = -.39$ ), with approval-dependent females tending to rate themselves as more improved. Ratings of own improvement correlated  $-.16$  for males and  $-.26$  for females with the number of hours of therapy.

To assess the stability of the index of need for approval during the course of therapy and as a further check on the legitimacy of administering the independent variable after termination to a sizeable proportion of the sample, the *M-C* scale was readministered to 24 patients who had been tested 5 months earlier and then remained in treatment. In this interval, patients completed between 20 and 25 hours of therapy. A test-retest correlation of .68 was obtained. The initial mean of the test-retest group was 10.92; at the second testing, the obtained mean was 10.42 ( $t = 0.57$ , *ns*). Moreover, in this subsample, the mean *M-C* scores are comparable to the mean of the remaining 61 patients (10.92 versus 12.25,  $t = 0.57$ ; 10.42 versus 12.25,  $t = 1.42$  *ns*). It remains to note that the correlation between the measure of need for approval and number of hours of therapy among the 34 patients in Group II<sub>a</sub> who took the scale at the Clinic was  $-.34$ . Partialing out the number of

hours of therapy at the time the scale was given, the correlation was  $-.29$ . This, of course, is almost identical to the  $-.28$  correlation reported in Table 3.

In an attempt to replicate the finding reported by Schachter (1959) that first-born and only-child patients remain longer in therapy than later borns, these two groups of patients were compared in length of stay in therapy. For 49 patients on whom ordinal position was recorded, a Mann-Whitney *U* test revealed no difference in hours of therapy ( $z = 0.22$ , *ns*).

### DISCUSSION

High and low need for approval patients differ strikingly in the length of time they remain in psychotherapy. Approval-dependent patients terminate much earlier than those less approval-motivated. To accept the hypothesis that the earlier termination of high need for approval patients represents a means of avoiding anticipated threats to self-esteem, it is first necessary to exclude the possibility that they are simply less disturbed and less in need of help than those less dependent on approval. As established by therapists' ratings, approval-dependent patients appear to be more defensive or disorganized. Moreover, they tend to be judged by their therapists as less personally liked, less satisfied with the progress of therapy, and less improved in treatment. Finally, no systematic differences were found between high and low need for approval groups in type or severity of diagnosis. It would appear that the approval-motivated group did not terminate earlier as a result of progress and improvement, and their early termination thus takes on the character of resistance and defensiveness.

These findings are consonant with recent studies on the approval motive (Conn, 1962; Crowne & Liverant, 1963) and support the conceptualization of the approval-dependent person as one who is concerned with protecting and maintaining a vulnerable self-image. Thus, it seems likely that, faced with the prospect of self-revelation and the threats this poses to his defensive image of himself, the approval-dependent patient convinces himself that psychotherapy is really not worth while or that, in a very few hours, he



has attained sufficient improvement. For the approval-dependent patient, the dilemma is one of giving up a defensive self-conception or defying the therapist. The outcome of such an avoidance-avoidance conflict is leaving the field. Additionally, it is possible that these patients fail to obtain highly valued affectional and dependency gratifications from their therapists which might alter the balance of the conflict. A corollary, but not incompatible interpretation, is that the approval-motivated individual is apprehensive about possible social criticism for seeking therapy and the (to him) implicit admission that he is "crazy." It is worth observing that the defensive test-taking behavior of high need for approval persons is consistent with this general interpretation of our major finding.

In this study, both time pressures and another large scale research project being conducted at the same time necessitated administering the *SD* scale to a large number of patients after the conclusion of psychotherapy. This raises a crucial methodological issue: whether, perhaps, scores on the scale were affected by participation in psychotherapy, thus rendering the findings more parsimoniously interpretable in terms of the influence of therapy on the questionnaire. As detailed above, an attempt was made to check on the possible effect of therapy on the scale by conducting a long term, test-retest analysis of the index of need for approval while patients were in treatment. No evidence for systematic changes was found. Also, the relationship with number of hours of therapy was maintained in the group given the *M-C* scale at the Clinic.

In fact, the test-retest data strongly suggest that the approval motive is very resistant to change. If any procedures could systematically alter dependence on the approval of others, it would certainly be expected that psychotherapy would be high on the list. Whether premature termination can be lessened among approval-dependent individuals is yet an open question, although it may be that, in the early hours of psychotherapy, before an enduring and trusting relationship with the therapist has been established, minimizing the demand for frank personal revelation and providing re-

assurance and support would tend to reduce this form of resistance.

Ratings of own improvement showed only a very slight relationship with hours of therapy: patients rating themselves as more improved tended to remain longer in therapy. While no relationship was found between the index of need for approval and own improvement ratings for males, approval-motivated females tended to rate themselves as more improved despite the earlier termination of the high need for approval group. The latter result lends further support to the defensiveness hypothesis.

The finding that therapists are able to make meaningful predictions and characterizations of their patients independent of the length of time the patient has been in therapy above a certain minimum, tends to support Meehl's (1960) demonstration that therapists arrive quite early at stable interpretations of their patients.

The discrepancies that were found between the therapists' ratings on the Case Rating scale and the ratings of the closing summaries raise some interesting issues. The Seeman scale ratings, in general, were the more predictively accurate; the terminal ratings, based on the therapists' impressions at the time of termination, did not correlate highly with hours of therapy, the index of need for approval, nor, surprisingly, the therapists' earlier ratings of improvement. Many therapists did not make closing summary statements on their patients; those who did gave accounts varying markedly in the amount of information conveyed. It seems clear that terminal statements of treatment are subject to vagaries that do not appear to affect in nearly the same degree structured ratings of patients.

In addition to their implications for the vulnerable self-esteem and defensiveness extension of the approval-motive construct, these results also have some suggestive value regarding the problem of a model for psychotherapy. The verbal conditioning paradigm has received considerable attention in the recent literature as an analogue of the process of psychotherapy (cf. Krasner, 1958). According to the verbal conditioning model, psychotherapy can be conceptualized

as a process in which the therapist subtly reinforces changes in the patient's verbal behavior; these changes then generalize to verbal and behavioral changes in the patient's real life outside of therapy. If we can assume that what is measured by the *M-C* scale is comparable in patients and college students, then the greater amenability to verbal reinforcement of approval-motivated individuals in a verbal conditioning situation (Crowne & Strickland, 1961; Marlowe, 1962) and their tendency to terminate therapy early and with less improvement pose a real problem for the model. Whatever the other merits or demerits of the verbal conditioning analogue, it certainly fails to account for the effect of individual differences and the meaning of the situation as it is perceived by the individual. It is one thing for the approval-dependent person to be more influenced by the subtle reinforcers of an experimenter to produce more plural nouns; it is evidently quite another thing to ask him to surrender his defensive conception of himself.

In summary, the results of this study suggest that defensiveness and avoidance of self-criticism constitute a major determinant of abrupt termination of psychotherapy. In contrast to previous studies, both social class and ordinal position were unrelated to length of psychotherapeutic stay.

## REFERENCES

- APFLECK, D. C., & MEDNICK, S. A. The use of the Rorschach test in the prediction of the abrupt terminator in individual psychotherapy. *J. consult. Psychol.*, 1959, 23, 125-128.
- AULD, F., & ERON, L. D. Use of Rorschach scores to predict whether patients will continue psychotherapy. *J. consult. Psychol.*, 1953, 17, 104-109.
- AULD, F., & MYERS, J. K. Contributions to a theory for selecting psychotherapy patients. *J. clin. Psychol.*, 1954, 10, 56-60.
- BARTHEL, C. E., & CROWNE, D. P. The need for approval, task categorization, and perceptual defense. *J. consult. Psychol.*, 1962, 26, 547-555.
- CONN, L. Instigation to aggression, emotional arousal, and defensive emulation. Unpublished master's thesis, Ohio State University, 1963.
- CROWNE, D. P., & LIVERANT, S. Conformity under varying conditions of personal commitment. *J. abnorm. soc. Psychol.*, in press.
- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, 24, 349-354.
- CROWNE, D. P., & STRICKLAND, BONNIE R. The conditioning of verbal behavior as a function of the need for social approval. *J. abnorm. soc. Psychol.*, 1961, 63, 395-401.
- FREUD, S. Further recommendations in the technique of psychoanalysis. On beginning the treatment. The question of the first communications. The dynamics of the cure. In, *Collected Papers*. Vol. 2. London: Hogarth Press, 1949.
- GALLAGHER, J. J. The problem of escaping clients in nondirective counseling. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. University Park, Pa.: Pennsylvania State Univ. Press, 1953.
- GARFIELD, S. L., & AFFLECK, D. C. Therapists' judgments concerning patients considered for psychotherapy. *J. consult. Psychol.*, 1961, 25, 505-509.
- GIBBY, R. G., STOTSKY, B. A., HILER, E. W., & MILLER, D. R. Validation of Rorschach criteria for predicting duration of therapy. *J. consult. Psychol.*, 1954, 18, 185-191.
- HILER, E. W. The sentence completion test as a predictor of continuation in psychotherapy. *J. consult. Psychol.*, 1959, 23, 544-549.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 55, 148-170.
- MARLOWE, D. Need for social approval and the operant conditioning of meaningful verbal behavior. *J. consult. Psychol.*, 1962, 26, 79-83.
- MARLOWE, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. consult. Psychol.*, 1961, 25, 109-115.
- MEERL, P. E. The cognitive activity of the clinician. *Amer. Psychologist*, 1960, 15, 19-27.
- ROSENTHAL, D., & FRANK, J. D. The fate of psychiatric clinic outpatients assigned to psychotherapy. *J. nerv. ment. Dis.*, 1958, 127, 330-343.
- SALMAN, A. R., & CROWNE, D. P. The need for approval, improvisation, and attitude change. Paper read at Midwestern Psychological Association, Chicago, May 1962.
- SCHACHTER, S. *The psychology of affiliation*. Stanford: Stanford Univ. Press, 1959.
- SEEMAN, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers and Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954.
- STRICKLAND, BONNIE R., & CROWNE, D. P. Conformity under conditions of simulated group pressure as a function of the need for social approval. *J. soc. Psychol.*, 1962, 58, 171-181.
- SULLIVAN, P. L., MILLER, CHRISTINE, & SMELSER, W. Factors of length of stay and progress in psychotherapy. *J. consult. Psychol.*, 1958, 22, 1-9.
- TAULBEE, E. S. Relationship between certain personality variables and continuation in psychotherapy. *J. consult. Psychol.*, 1958, 22, 83-89.

(Received July 3, 1962)



## INTELLIGENCE TEST PERFORMANCES OF BRAIN DAMAGED SUBJECTS WITH LATERALIZED MOTOR DEFICITS<sup>1</sup>

HOMER B. C. REED, JR., AND R. M. REITAN<sup>2</sup>

*Indiana University Medical Center*

The Wechsler-Bellevue Intelligence Test was administered to groups of brain damaged Ss having lateralized motor deficits, brain damaged Ss with no lateralized motor deficits, and control Ss having no brain damage. The results indicated that brain damaged Ss with right-sided motor deficits demonstrate different kinds of intellectual impairment than brain damaged Ss with left motor deficits and, additionally, that the intellectual impairment is a function of the lateralization of cerebral dysfunction rather than reflecting the motor handicap per se. The significance of the findings was briefly discussed.

Since 1950 a variety of investigative efforts by psychologists and neurologists have established that certain basically different intellectual abilities are impaired in individuals with brain damage confined principally to the left cerebral hemisphere as compared with individuals in whom brain damage is confined principally to the right cerebral hemisphere. In cases of hemiplegia or hemiparesis, where there is adequate independent evidence of cerebral damage, one might therefore expect to find differential impairment of intellectual functions attributable to organic involvement principally of either the left or right cerebral hemisphere. Additionally, such cases might demonstrate deficiencies on measures of intelligence attributable primarily to the motor involvement. Such deficiencies would presumably be greatest on tasks involving complex psychomotor functions and should parallel in magnitude the degree of severity of the motor handicap.

The present study has been concerned with comparing the intelligence test performances of groups of brain damaged subjects with lateralized motor deficits, a group of brain damaged subjects without lateralized motor deficits and a group of non-brain-damaged control subjects. The intelligence test employed was the Wechsler-Bellevue (Form 1). This test consists of 11 sub-

tests, 6 of which were designed to measure verbal abilities, with the remaining 5 designed to measure nonverbal abilities. Differential impairment on this test of the performances of the various experimental groups would be an objective demonstration that brain damaged subjects with left-sided motor deficits may be expected to show different kinds of intellectual deficits than brain damaged subjects with right-sided motor deficits. Such a finding would constitute additional evidence that the left and right cerebral hemispheres subserve different intellectual functions. Furthermore, the design of the present study permits objective evaluation of the influence of centrally determined lateralized motor deficits on measures of intelligence involving complex psychomotor functions.

### METHOD

The following groups of subjects, used in this study, were selected from individuals tested at the Neuropsychology Laboratory of the Indiana University Medical Center during the period of 1951-60.

- Group I—18 brain damaged subjects with right hemiplegia
- Group II—20 brain damaged subjects with right hemiparesis
- Group III—20 brain damaged subjects with left hemiplegia
- Group IV—20 brain damaged subjects with left hemiparesis
- Group V—20 brain damaged subjects without lateralized motor deficits
- Group VI—20 control subjects without cerebral damage

<sup>1</sup>Supported by Research Grant B-2416 from the National Institute of Neurological Diseases and Blindness, Public Health Service.

<sup>2</sup>The authors are indebted to Melinda Mayhall for assistance with the statistical analysis.

TABLE 1  
AGE AND EDUCATION VARIABLES  
FOR EACH GROUP

Group	Mean age	Range	Mean education	Range
I	47.00	21-61	10.94	6-16
II	44.80	26-63	9.50	7-16
III	42.25	29-55	9.95	2-19
IV	43.20	20-65	9.90	3-16
V	44.60	17-64	9.95	3-16
VI	44.70	27-62	10.15	3-18

Note.—Group I, Brain damaged patients with right hemiplegia; Group II, Brain damaged patients with right hemiparesis; Group III, Brain damaged patients with left hemiplegia; Group IV, Brain damaged patients with left hemiparesis; Group V, Brain damaged patients with no lateralized motor deficits; Group VI, Controls without cerebral damage.

All groups were matched as closely as possible on the variables of age and educational levels and there were no significant intergroup differences on these variables. The brain damaged groups were matched as closely as possible with respect to type of lesion, but as indicated in Table 2 it was not possible to obtain an exact match on this variable. The diagnosis of brain damage was based on a complete neurological examination and on such additional procedures as were deemed appropriate for each subject (i.e., EEG, angiography, pneumoencephalography, neurosurgical findings, etc.). The criteria for the diagnosis of hemiplegia or hemiparesis included, for each subject, the notes of the examining neurologists or neurosurgeons made during the subject's course of hospitalization, plus additional confirmatory test findings obtained in the Neuropsychology Laboratory on Halstead's Finger Oscillation Test and Halstead's (1947) Tactual Performance Test. It should be noted that although none of the brain damaged subjects in Group V were judged to have lateralized motor deficits, the

TABLE 2  
DIAGNOSTIC DISTRIBUTION OF PATIENTS

Diagnosis	Group				
	I	II	III	IV	V
Tumor					
Cerebral vascular disease	2	7	4	5	5 <sup>a</sup>
Degenerative or demyelinating disease	9	7	11	9	10
Infectious or inflammatory disease	2	1	1	1	1
Trauma	1	1	0	0	0
	4	4	4	5	4

Note.—Group I, Brain damaged patients with right hemiplegia; Group II, Brain damaged patients with right hemiparesis; Group III, Brain damaged patients with left hemiplegia; Group IV, Brain damaged patients with left hemiparesis; Group V, Brain damaged patients with no lateralized motor deficits; Group VI, Controls without cerebral damage. Includes two extrinsics.

criteria did not preclude the presence of motor deficits approximately equilateral in nature.

The control group included subjects hospitalized for conditions not relating to brain dysfunction and nonhospitalized subjects.

After the Wechsler-Bellevue Test had been administered to each subject, all scores for each subtest and for the summary measures were pooled and ranked within each distribution from poorest to best performance. The ranks were then converted to normalized *T* score distributions, giving each distribution a mean of 50 with a standard deviation of 10. Differences between the groups of subjects were evaluated by computing the appropriate *t* ratios.

## RESULTS

Table 3 presents the performances of all groups of subjects on the Wechsler variables. The right hemiplegia group performed at a slightly higher level than the right hemiparetic group on 10 of the 11 Wechsler subtests. Comparison of the summary measures on these two groups, however, reveals no significant differences (see Table 4). Subjects with left hemiplegia were superior to subjects with left hemiparesis on 5 of the 11 subtests and again the comparison of the summary measures reveals no significant differences. It is readily apparent, then, that the severity dimension did not reliably order the psychological test results. In the absence of any significant differences between the hemiplegic and the corresponding hemiparetic groups it was decided to combine the four groups with lateralized motor deficits into a right motor deficit group and a left motor deficit group. Figure 1 presents the performances of the combined groups on the Wechsler subtests. As can be seen, the normal controls were superior to all of the brain damaged groups on all subtests. The group with left-sided motor deficits was superior to the group with right motor deficits on all of the verbal measures with the reverse being true on all of the performance measures. The brain damaged group having no lateralized motor deficit performed at a level between the other brain damaged groups on the verbal tests but was generally somewhat superior to them on the performance measures. This last finding suggests that subjects having lateralized motor deficits may be at a slight disadvantage on



TABLE 3  
T SCORE MEANS AND STANDARD DEVIATIONS OF WECHSLER VARIABLES FOR ALL GROUPS

Variable	Group					
	I	II	III	IV	V	VI
Verbal WS						
$\bar{X}$	48.33	43.65	51.30	50.50	47.15	58.95
$SD$	7.57	10.20	9.56	8.96	9.03	7.95
Performance WS						
$\bar{X}$	49.78	49.05	44.30	46.70	50.45	59.80
$SD$	6.47	8.32	9.08	10.65	10.17	7.64
Information						
$\bar{X}$	49.50	45.15	52.65	49.20	46.15	57.35
$SD$	9.71	11.27	8.95	8.09	8.38	9.35
Comprehension						
$\bar{X}$	47.61	44.60	51.65	49.90	49.25	56.25
$SD$	9.36	10.79	11.14	7.55	9.54	7.40
Digit span						
$\bar{X}$	48.00	44.65	50.70	51.55	48.40	57.00
$SD$	7.70	7.56	10.72	7.97	8.89	9.06
Arithmetic						
$\bar{X}$	49.61	44.15	49.65	51.35	47.85	57.35
$SD$	7.76	10.00	9.63	8.91	9.68	6.78
Similarities						
$\bar{X}$	48.61	44.75	51.90	49.05	47.90	57.45
$SD$	8.54	9.67	8.83	9.08	11.04	7.74
Vocabulary						
$\bar{X}$	48.72	46.00	51.20	50.90	48.55	54.75
$SD$	10.35	10.99	9.91	7.79	8.52	10.60
Picture arrangement						
$\bar{X}$	51.78	51.05	45.10	46.85	49.50	56.45
$SD$	5.44	9.43	8.63	9.90	11.05	9.57
Picture completion						
$\bar{X}$	51.94	47.35	46.80	45.95	50.15	57.80
$SD$	8.12	7.88	9.59	8.29	12.24	6.44
Block design						
$\bar{X}$	49.89	48.90	42.70	47.40	52.55	59.00
$SD$	5.73	8.46	9.21	9.22	9.94	7.05
Object assembly						
$\bar{X}$	50.00	50.30	43.50	48.40	51.80	57.05
$SD$	9.03	7.99	8.14	10.24	10.47	7.35
Digit symbol						
$\bar{X}$	48.94	47.20	46.30	48.70	49.75	59.15
$SD$	7.33	7.82	9.49	10.40	9.54	7.73

Note.—Group I, Brain damaged patients with right hemiplegia; Group II, Brain damaged patients with left hemiplegia; Group III, Brain damaged patients with left hemiplegia; Group IV, Brain damaged patients with right hemiplegia; Group V, Brain damaged patients with no lateralized motor deficits; Group VI, Controls without cerebral damage.

complex psychomotor problems, but the handicap does not seem to be a significant one.

Table 4 presents statistical comparisons on the Wechsler summary measures of Verbal Weighted Score and Performance Weighted Score for the various groups of subjects.

The controls were significantly superior to the brain damaged subjects on both of these measures. The left motor deficit group was significantly superior to the right motor deficit group on Verbal Weighted Score. In spite of the consistency, however, with which the right motor deficit group excelled the

TABLE 4  
I RATIOS ON WECHSLER WEIGHTED SCORES

Group	VWS		PWS		VWS - PWS	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
I versus II	1.59	.20	.30	.80		
III versus IV	.27	.80	.77	.50		
I + II versus III + IV	2.41	.02	1.97	.10	4.64	.001
V versus I + II	.51	.70	.45	.70	.10	1.00
V versus III + IV	1.50	.20	1.82	.10	3.96	.001
V versus I + II + III + IV	.55	.60	1.33	.20		
VI versus I + II	5.37	.001	5.03	.001	1.13	.30
VI versus III + IV	3.35	.005	5.69	.001	2.93	.01
VI versus V	4.39	.001	3.29	.005	1.03	.40

Note.—Group I, Brain damaged patients with right hemiplegia; Group II, Brain damaged patients with right hemiparesis; Group III, Brain damaged patients with left hemiplegia; Group IV, Brain damaged patients with left hemiparesis; Group V, Brain damaged patients with no lateralized motor deficits; Group VI, Controls without cerebral damage.

left motor deficit group on the performance subtests, the difference between these two groups on the Performance Weighted Score did not reach statistical significance. The groups of brain damaged subjects with

lateralized motor deficits did not differ significantly from the group of brain damaged subjects without lateralized motor deficits on either summary measure.

Evaluation of intraindividual difference

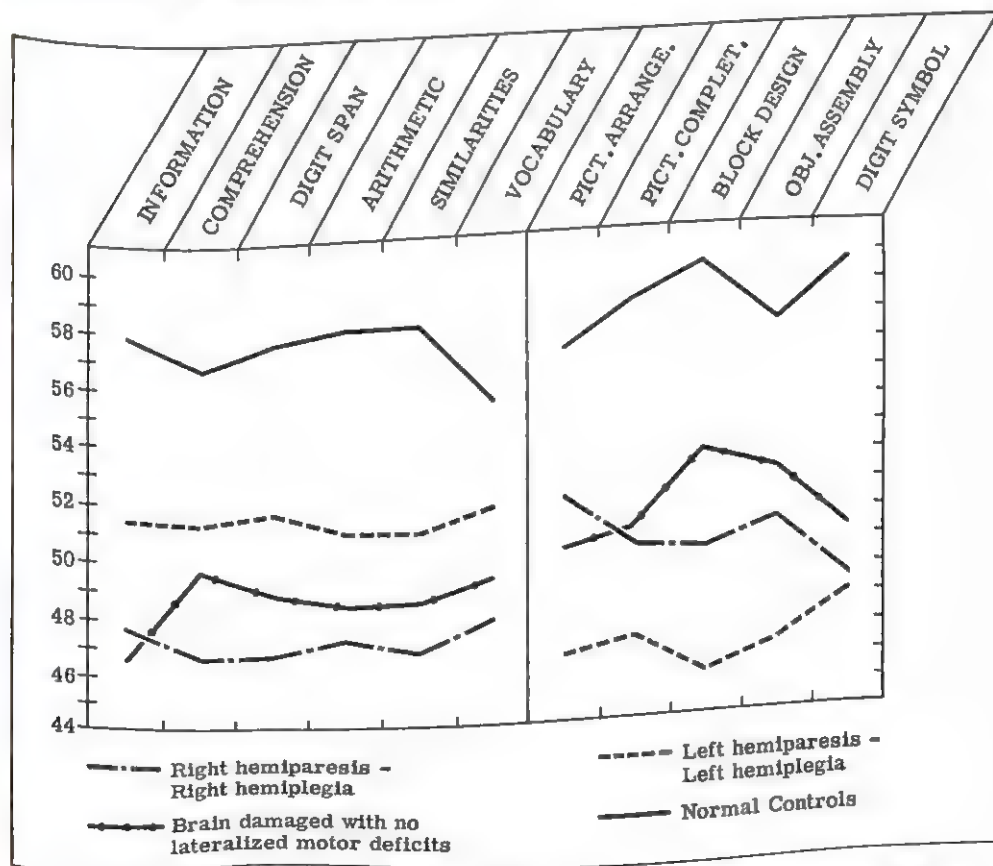


FIG. 1. Comparative Wechsler Bellevue performances for brain damaged groups with and without lateralized motor deficits and for a control group without brain damage.



scores between the Verbal and Performance Weighted Scores reveals a clear and significant difference between the right and the left motor deficit groups. The group of subjects with left motor deficits also differed significantly from the brain damaged and normal control groups on this measure, but the right motor deficit group did not.

#### CONCLUSIONS AND DISCUSSION

The principal conclusions to be drawn from this study may be stated as follows: brain damaged subjects with right motor deficits may be expected to demonstrate a different kind of intellectual impairment than brain damaged subjects with left motor deficits. The differential impairment of these groups of subjects is attributable to lateralization of cerebral dysfunction and seems to be unrelated to lateralization of the motor deficit per se. These findings support the research of Briggs (1960), who found that subjects performed nearly as well with either hand as with both hands on the performance subtests of the Wechsler Adult Intelligence Scale. The conclusions of the present study are also in close agreement with the researches of Reitan (1955), Kløve (1959), and Doehring, Reitan and Kløve (1961), who have demonstrated that impairment of verbal abilities as measured by the Wechsler-Bellevue Scale is associated primarily with lesions of the left cerebral hemisphere, whereas abilities requiring the appreciation of spatial and temporal relationships are impaired primarily by lesions of the right cerebral hemisphere. Each of the above studies employed different lateralizing criteria but each used subjects with current neurological complaints, as has the present investigation. Thus the conclusion relating different abilities to the two cerebral hemispheres is seen to have a high degree of generality for this population of patients. Failure to find differential intellectual impairment in groups with longstanding damage having lateralizing neurological signs has been reported by Fitzhugh, Fitzhugh, and Reitan (in press) and Kløve and Fitzhugh (in press), suggesting that chronicity of damage may be an important independent variable. Bruell and his associates (Bruell, Peszczynski, & Albee,

1956; Bruell, Peszczynski, & Volk, 1957) found no differences in ability to perceive verticality among groups of brain damaged subjects with differently lateralized motor deficits. This research, also, suggests the possible importance of chronicity of cerebral dysfunction as a determinant of intellectual impairment.

The presence of conflicting findings concerning the existence of differential intellectual impairment in brain damaged subjects having lateralizing neurological signs emphasizes the complexity of the criterion problem in this area of research. The validity with which neurological signs indicate the organic condition of the brain is dependent upon a wide variety of covarying factors and until this variance is systematically explored, conflict rather than agreement among the results of various investigations may be expected to occur.

#### REFERENCES

- BRIGGS, P. F. The validity of WAIS performance subtests completed with one hand. *J. clin. Psychol.*, 1960, 16, 318-320.
- BRUELL, J. H., PESZCZYNSKI, M., & ALBEE, G. W. Disturbance of perception of verticality in patients with hemiplegia: A preliminary report. *Arch. phys. Med. Rehabil.*, 1956, 37, 677-681.
- BRUELL, J. H., PESZCZYNSKI, M., & VOLK, D. Disturbance of perception of verticality in patients with hemiplegia: Second report. *Arch. phys. Med. Rehabil.*, 1957, 38, 776-780.
- DOEHRING, D., REITAN, R. M., & KLØVE, H. Changes in patterns of intelligence test performances associated with homonymous visual field defects. *J. nerv. ment. Dis.*, 1961, 132, 227-233.
- FITZHUGH, KATHLEEN B., FITZHUGH, L., & REITAN, R. M. Wechsler-Bellevue comparisons in groups with "chronic" and "current" lateralized and diffuse brain lesions. *J. consult. Psychol.*, in press.
- HALSTEAD, W. C. *Brain and intelligence: A quantitative study of the frontal lobes*. Chicago, Ill.: Univ. Chicago Press, 1947.
- KLØVE, H. Relationship of sensory imperception to distribution of Wechsler-Bellevue scores. Unpublished manuscript, Indiana University Medical Center, 1959.
- KLØVE, H., & FITZHUGH, KATHLEEN B. The relationship of differential EEG patterns to the distribution of Wechsler-Bellevue scores in a chronic epileptic population. *J. clin. Psychol.*, in press.
- REITAN, R. M. Certain differential effects of left and right cerebral lesions in human adults. *J. comp. physiol. Psychol.*, 1955, 48, 474-477.

(Received June 19, 1962)

## SOCIAL DESIRABILITY UNDER ROLE PLAYING INSTRUCTIONS:

### A REPLY TO WALKER

JERRY S. WIGGINS

*University of Illinois*

Wiggins (1959) found that Edwards' SD scale was not effective in distinguishing MMPI records of Ss instructed to answer in terms of social desirability from records of other Ss not so instructed. Walker (1962) has criticized Wiggins' study on both procedural and interpretative grounds and reported findings of a redesigned replication that suggest SD can perform this discrimination. It was argued that given (a) the actual development of the SD scale, (b) the success of other desirability scales in both studies, (c) the similarity of control scales in both studies, and (d) Walker's failure to control for the test-retest "improvement" effect, Walker's conclusions cannot stand, and his defense of the SD scale must be judged unsuccessful.

In an earlier experiment, Wiggins (1959) presented evidence which suggested, among other things, that the social desirability scale (SD) of Edwards (1957) was not effective in distinguishing MMPI records of subjects instructed to answer in terms of social desirability from records of subjects not so instructed. More recently, Walker (1962) has criticized Wiggins' study on both procedural and interpretative grounds and reported findings of a redesigned replication that reflect more favorably on Edwards' SD scale as a discrimination index.

Wiggins (1959) employed two randomly constituted groups: (a) a control group of 190 college students receiving the standard instructions printed on the front of the MMPI booklet; and (b) an experimental group of 250 college students instructed to answer items in the direction that people in general would consider to be more desirable. With respect to the experimental group, Wiggins (1959) assumed that:

... a relatively uniform group of high social desirability responders may be created by special instructions to the subjects. Such a group would differ in the amount of this tendency and could, in principle, serve as a criterion group for distinguishing an unusual amount of social desirability response style from a small amount. Scales which purport to measure social desirability response style should presumably achieve a certain amount of success in differentiating such a group from a "normal" group (p. 420).

Walker (1962) objected to Wiggins' procedure on three grounds: (a) Edwards' SD scale was developed to measure social desirability response tendencies under *standard* test conditions and hence should not be required to discriminate between standard and role playing situations; (b) social desirability was probably operative under Wiggins' control conditions and this, together with the "obvious" nature of SD items, would contribute to a high mean scale score that would be restricted in the amount it could shift under role playing instructions; and (c) Wiggins' instructions or his subjects' ability to grasp them may have been less than optimal from the standpoint of maximizing the amount of social desirability responding in the experimental group.

With the above objections in mind, Walker conducted an experiment in which the same 120 subjects were given standard instructions followed, 1 week later, by experimental instructions designed to "closely approximate those used in the original development of Edwards' SD scale" (Walker, 1962, p. 163). The marked improvement in the discriminative efficiency of the SD and K scales under Walker's revised procedures led him to conclude that his three criticisms of Wiggins' study (as it related to Edwards' SD scale) were appropriate.

The present paper will endeavor to demonstrate that given (a) the way in which



TABLE 1

COMPARISON OF THREE SOCIAL DESIRABILITY ROLE PLAYING INSTRUCTIONS

Wiggins (1959)	Walker (1962)	Edwards (1958) <sup>a</sup>
<p>This inventory consists of numbered statements. Read each statement and decide whether <i>People in General</i> would consider a true or a false answer to be more desirable. You are <i>not</i> asked whether the statement is true or false as applied to you. Rather you are asked to decide which answer you think <i>People in General</i> would consider to be more desirable</p> <p>You are to mark your answers on the answer sheet you have. Look at the example on the answer sheet shown at the right. If you decide that people in general consider true to be the most desirable answer, blacken between the lines in the column headed T. (See A at right.) If you decide that people in general consider false to be the most desirable answer, blacken between the lines in the column headed F. (See B at right.)</p> <p>Remember <i>do not give your own opinion of yourself</i> but blacken the answer that you feel people in general would consider to be most desirable.</p> <p>In marking your answers on the answer sheet, <i>be sure that the number of the statement agrees with the number on the answer sheet.</i> Make marks heavy and black. Erase completely any answer you wish to change. Do not make any marks on this booklet.</p> <p>Be sure to answer <i>every</i> statement.</p>	<p>Last time when we gave you the MMPI, we asked you to respond to the items honestly and to tell us whether each item was True or False as applied to you. To-day, we want to see how well you can do in answering the same items so as to create the most favorable impression you can. In other words, we want you to answer each item in terms of whether you consider the True or the False response to be more desirable. If you think the True response is more desirable than the False then answer the item True. If you think the False response is more desirable than the True, then answer the item False. For example, if you were trying to create a favorable impression, you would probably answer True to this item:</p> <p>I am regarded as an honest person.</p> <p>and you would probably answer False to this item:</p> <p>I frequently hurt the feelings of others.</p> <p>Remember, answer each item in such a way as to create the most favorable or desirable impression possible.</p>	<p>Please fill this out for Dr. Edwards. He wishes to obtain social desirability ratings for these items. Mark an item T (true) if marking it true is the most socially desirable response to the item. Mark an item F (false) if marking it false gives the most socially desirable response. Give the most socially desirable picture you possibly can from the way you mark the items. Please mark every item. . .</p> <p>No signatures are needed.</p>

<sup>a</sup> A. L. Edwards, personal communication, February 20, 1958.

Edwards in fact developed the *SD* scale, (b) the discriminatory power of other desirability scales in both the Wiggins and Walker studies, (c) the similarity of control scale scores in both studies, and (d) Walker's failure to correct for the spurious test-retest "improvement" phenomenon on the MMPI, Walker's conclusions cannot stand, and his defense of the Edwards' *SD* scale must be judged unsuccessful.

#### ROLE PLAYING INSTRUCTIONS AND SOCIAL DESIRABILITY SCALES

In criticizing the original experiment, Walker impeaches the instructions used as well as the ability of the subjects to grasp them: "... there is reason to believe that Wiggins' (1959) experimental instructions were such that they did not maximize the amount of responding in the socially desirable way" (p. 162); and "Wiggins' failure to demonstrate substantial differences in the *K* and *SD* scales may have been due to the failure of his subjects to understand fully the experimental instructions" (p. 163).

A comparison of the two sets of instructions (Table 1) suggests that Walker's objections are based on the undistinguished performance of *SD* and *K* in Wiggins' study rather than on any notable omission in instructions subsequently remedied by Walker. The implication that Stanford University undergraduates (who participated in the original study) are less astute in grasping the requirements of a role playing task than their counterparts at Western Washington College of Education may be somewhat reluctantly conceded as a possibility. However, one cannot help but admire the manner in which the Stanford students applied the instructions to items from the *Sd* and *Cof* pools,<sup>1</sup> even though they stumbled on items from Edwards' *SD*. The fact that results with Cofer's scale, *Cof*, replicated those of a study performed a decade before on a more heterogeneous population (Cofer, Chance, & Judson, 1949) can be cited as additional evidence of the representative-

ness of the Stanford students. Not least, the splendid success of Wiggins' *Sd* in Walker's own data likewise attests to this scale's indifference to changes of collegiate environment.

#### CONSTRUCTION OF EDWARDS' *SD* SCALE

Walker's "primary criticism" of Wiggins' (1959) experiment alleges a failure to recognize Edwards' (1961) position that the *SD* scale: "provides a measure of the tendency of subjects to give socially desirable responses in self-description under the *standard* instructions ordinarily used with personality inventories" (p. 353, *italics* his). Despite the fact that more than a dozen articles and a substantial portion of a book have been devoted to various aspects of Edwards' *SD* scale, the original instructions to the 10 judges<sup>2</sup> from whose responses the scale was developed have never been published. These instructions are presented in the third column of Table 1. Those familiar with the extensive literature concerning the *SD* scale, but unfamiliar with details of its construction, may be surprised to note that rather than being developed with reference to social desirability scale values of items, *SD* items were keyed on the basis of dichotomous responses of a small group of *role playing* subjects. Moreover, Edwards' *SD* scale was developed by use of role playing instructions not only similar to Walker's but very much like those employed in the development of Wiggins' role playing scale, *Sd* (Table 1). It is not clear why one scale (*Sd*) is held to be appropriately evaluated by role playing experiments while the other (*SD*) is not.

In constructing *Sd*, Wiggins (1959) employed the empirical method of *contrasted groups* in which an item analysis selected those items that significantly differentiated responses of a role playing group from responses of a standard group.<sup>3</sup> Edwards' *SD* scale was constructed solely from role playing

<sup>2</sup> Two secretaries, several graduate students, one staff member, and several undergraduates.

<sup>3</sup> The use of role playing designs for evaluating "faking good" tendencies has a history that considerably antedates Edwards' (1957) recent reformulation of the issue of "social desirability" (see, for example, the studies of Cofer, Chance, & Judson, 1949; Gough, 1952; Meehl & Hathaway, 1946; Ruch, 1942).

<sup>1</sup> *Sd* is a social desirability scale developed from items that discriminated experimentals from controls in subgroups of Wiggins' (1959) original study. *Cof* is a similar scale, developed from role playing groups (Cofer, Chance, & Judson, 1949).



responses and hence, the significance of a high *SD* score must be determined from further empirical investigation. Edwards (1957) assumed that subjects who score high on *SD* under standard instructions are behaving in a manner similar to that of his role playing judges. This is probably true, but if the behavior of such subjects cannot be distinguished from that of the rest of the standard group or from any other reasonable comparative group, then the fact that their responses resemble those of role players is without discriminative significance (Wiggins, 1962). Available item frequency data suggest that, on the average, items in Edwards' *SD* scale will be answered in the desirable direction by approximately 80% of any normative group. The argument that we will never know to what extent a "normative" group is responding in terms of social desirability only highlights the need for operational definitions that have independent empirical significance.

Walker (1962) evinces partial recognition of the foregoing state of affairs in noting that: "... the items of the *SD* scale are fairly obvious in terms of social desirability-undesirability. It seems reasonable that such would tend to elevate the control mean of the *SD* scale" (p. 162). However, this fact must be attributed to the items in the *SD* scale and not (as Walker does) to Wiggins' procedure, since the mean *SD* score in Wiggins' control group (30.9) is virtually identical to the mean *SD* score in Walker's control group (30.5).

#### REPEATED MEASURES VERSUS INDEPENDENT GROUPS DESIGNS

The most obvious procedural difference between the experiments of Wiggins (1959) and Walker (1962) is in the latter's use of a design in which the same subjects served as both role players and controls. The major rationale behind the use of independent contrasted groups was given by Wiggins (1959) in the original article: "... the population about which we would generalize contains subgroups of high dissimulators and relatively honest people—rather than people taking the MMPI under several conditions" (p. 240).

In addition to providing a more realistic basis for generalization, the contrasted groups

design avoids the necessity of employing a test-retest control group for the "improved adjustment" effect that frequently occurs with repeated administrations of the same inventory (Windle, 1954, 1955). When dissimulation of social desirability is called for on the second testing, improved adjustment effects operate in the direction of the hypothesis and an additional control group which simply takes the test twice under standard conditions would seem to be appropriate.

Windle (1954) has reviewed the rather extensive evidence for an improved adjustment effect in repeated administrations of personality questionnaires. In a later study specifically directed toward the MMPI, Windle (1955) administered the inventory twice to standardly instructed undergraduates, with a 1-week interval between testings. His conclusions are particularly germane to Walker's procedure:

Confirmation is gained from this study for the earlier finding that when personality tests are given a second time after a short time interval, subjects tend to get better adjustment scores than on the original testing. With regard to the MMPI, this retest effect seems greatest on the so-called "anxiety" scales. The validity scales also tend to reflect this retest effect (p. 248).

The Taylor *MA* and *K* scales both showed significant ( $p < .01$ ) improvement effects in Windle's study. Walker's demonstration of improvement in *SD* (which consists of 56% *MA* items and 13% *K* items) and in the *K* scale is therefore difficult to interpret in the absence of an appropriate control group.

By use of a repeated measures design, Walker was able to reduce the false negative identification rate of *SD* in his experimental group by some 30% over that found in Wiggins' experimental group. From a practical standpoint it is interesting to note that this improvement was accomplished by an overall increase of experimental group *SD* scores of only three points over that of Wiggins' experimental group. From a scientific standpoint, it would be interesting to know if this three-point increment might be observed in a control test-retest group that took the inventory twice under standard instructions. Windle's (1955) findings suggest this might be the case.

## REFERENCES

- COFER, C. N., CHANCE, JUNE, & JUDSON, A. J. A study of malingering on the MMPI. *J. Psychol.*, 1949, 27, 491-499.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden Press, 1957.
- EDWARDS, A. L. Social desirability or acquiescence in the MMPI? A case study with the *SD* scale. *J. abnorm. soc. Psychol.*, 1961, 63, 351-359.
- GOUGH, H. G. On making a good impression. *J. educ. Res.*, 1952, 46, 33-42.
- MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, 30, 525-564.
- RUCH, F. L. A technique for detecting attempts to fake performance on a self-inventory type of personality test. In Q. McNemar & Maud A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942.
- WALKER, J. N. An examination of the role of the experimentally determined response set in evaluating Edwards' Social Desirability Scale. *J. consult. Psychol.*, 1962, 26, 162-166.
- WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.*, 1959, 23, 419-427.
- WIGGINS, J. S. Strategic, method and stylistic variance in the MMPI. *Psychol. Bull.*, 1962, 59, 224-242.
- WINDLE, C. Test-retest effect on personality questionnaires. *Educ. psychol. Measmt.*, 1954, 14, 617-633.
- WINDLE, C. Further studies of test-retest effect on personality questionnaires. *Educ. psychol. Measmt.*, 1955, 15, 246-253.

(Received June 25, 1962)



## MMPI DIFFERENCES ASSOCIATED WITH SEX, RACE, AND CLASS IN TWO ADOLESCENT SAMPLES<sup>1</sup>

ROBERT L. McDONALD

AND

MALCOLM D. GYNTHIER

*Medical College of South Carolina*

*Washington University Medical School*

This study was designed to determine if Negroes and whites perform differently on the MMPI with socioeconomic status controlled. The sample of 360 was drawn from high school seniors attending segregated southern schools. Class categorization was based on parental occupation. Analyses of variance and supplementary "t" tests showed that Negro students obtained higher scores than whites on Scales L, F, 1, 2, 5, 8, and 9 and lower scores on Scale 3, and males obtained higher scores than females on Scales 1, 2, 5, 7, and 9 and lower scores on 0, but that no differences were attributable to the class factor. It was hypothesized that the Negro-white differences reflect distinctive cultural patterns. The development of separate MMPI norms for Negroes was recommended.

*An MMPI Handbook* (Dahlstrom & Welsh, 1960) indicates that only four studies comparing the performance of Negroes and whites on this instrument were carried out during the period 1939-60. Three of these studies were of prisoners in state institutions (Caldwell, 1954; Fry, 1949; Panton, 1959); the other was of patients in a Veterans Administration hospital (Hokanson & Calden, 1960). The Negro samples generally obtained higher scores on the clinical scales and were significantly higher on Scales 8 (*Sc*) and 9 (*Ma*) in three of the four studies cited. It is difficult, however, to generalize from these findings as the samples were drawn from populations clearly not representative of Negroes-in-general.

A review of the literature to date discloses that only two additional studies of this problem have been done (Ball, 1960; Miller, Wertz, & Counts, 1961). The latter experiment used Veterans Administration mental hygiene clinic patients which, again, could not be considered as a representative sample of Negroes. Ball used high school students, but his sample was so small—14 Negro males and 17 Negro females—that his findings of a high incidence of neurotic tendencies among the Negro boys as well as withdrawal and

introversion among the Negro girls must be considered as highly tentative.

The relationship between socioeconomic status and MMPI performance has been investigated by Gough (1948), Nelson (1952), and Perlman (1950). The major finding was that *K* scores are significantly positively correlated with various status indices. Also, lower-class subjects generally obtained higher scores on the clinical scales, although the results showed few specific agreements. That is, Nelson (1952), using psychiatric patients, found that lower-class subjects got significantly higher scores on Scales 1, 3, and 6, whereas Perlman's (1950) significant differences with normal subjects were on Scales 1, 7, and 8. Furthermore, the latter investigator found no status effects when he examined the MMPIs of working-class and middle-class psychiatric patients. Dahlstrom and Welsh (1960) interpret these data as indicating that middle-class and upper-class test subjects are more cautious about saying things that may adversely affect their reputations and that lower-class subjects approach the MMPI with less desire to cover up and erect a facade of adequacy. Their conclusions appear to be principally based on the positive *K*-status relationship and imply that lower-class subjects would obtain significantly lower *K* scores than upper-class subjects.

Dahlstrom and Welsh (1960) in the *Handbook* relate the race and status research data by stating that "the findings obtained in all studies except that of Fry are the sort

<sup>1</sup> The authors express appreciation to Thomas A. Carrere, Eugene C. Clark, Miss Margaret Broadnax, Jacob O. Sines, and Joseph L. White for their invaluable assistance during various stages of the study.

that would be expected from known effects of socioeconomic inequities." The purpose of the present study is to test Dahlstrom and Welsh's hypothesis, using samples which are representative of the adolescent subgroup of the racial group and which are large enough to yield stable results.

### METHOD

This experiment utilized a  $2 \times 2 \times 3$  factorial design. Race, sex, and socioeconomic level were the factors controlled. Three hundred and sixty subjects were included in the study, with 30 subjects in each of the 12 cells. This sample was drawn from a larger group which consisted of 354 (196 female and 158 male) Negro and 263 (132 female and 131 male) white high school seniors who comprised consecutive graduating classes (1961 and 1962) of urban segregated Southern high schools. Ages ranged from 16 to 19 with no mean age differences between sex or race.

An occupational classification scheme by Schneider and Lysgaard (1953) was applied to subjects' parents in order to classify subjects with regard to the social class factor. Four occupational classes are derived in terms of the degree of supervisory power over "lower" occupations and independence of supervisory control from "higher" occupations. The four classes are:

1. Independent occupations (including executives, directors, owners of business or farm, doctors, lawyers, bankers, ministers, professors, consulting engineers)
2. Dependent occupations involving skill and supervision or manipulation of others (including supervisors, foremen, technical engineers in industrial employment, sales workers, agents, public officials, entertainers)
3. Dependent occupations involving skill but little supervision or manipulation of others (including skilled and semiskilled workers and workers in industrial employment)
4. Dependent occupations involving little skill and little supervision or manipulation of others (including assembly line workers, laborers, janitors, farm workers, road workers, miners, drivers)

In addition, the present study employed an "unclassifiable" category for those parents listed as deceased, living in another city with job unspecified, or as unemployed. The groupings were made, when possible, on the basis of paternal occupation. In that minority of cases where the father's occupation was unclassifiable, the maternal occupational status was used for classification purposes. It should be pointed out that the Schneider-Lysgaard classification scheme is relatively crude. The use of additional indices such as the Sims (1927) Score Card should lead to more precise control of the status factor.

MMPIs were administered by school counselors during special group testing periods. Each student

was also instructed to list the occupational status and job title, if known, of both parents on a form provided for this purpose. Inspection of the MMPI answer sheets revealed no invalid records either through inability to understand the instructions or through omission of an excessive number of items. Furthermore, MMPIs were not discarded because of high *L* or *F* scores as several studies (e.g., Gynther, 1961) have suggested that such configurations are not invalid, but meaningfully related to personality traits such as hostility.

Two judges, the junior author and a second clinical psychologist, independently rated the maternal and paternal occupations in terms of the various classifications. In case of disagreement between these raters, the final rating was made by the senior author. Inspection of the socioeconomic classifications revealed skewed distributions for both races with a disproportionate number of the white sample falling within the upper socioeconomic level while the reverse applied to the Negroes. To insure sufficient numbers in each cell, it was necessary to combine Classes 1 and 2 which resulted in the use of only three socioeconomic categories in the analysis.

The total group of 617 was reduced to the experimental sample of 360 by alphabetizing the protocols within each of the 12 subgroups and deleting subjects by randomization procedures until the subgroups were all of a size equal to the smallest subgroup (i.e., 30).

### RESULTS

Table 1 shows cell means and standard deviations for the 13 MMPI scales. Tests for homogeneity of variance (Dixon & Massey, 1951) resulted in rejection of the hypothesis of homogeneity for Scales 1 (*Hs*) and 3 (*Hy*). Table 2 contains the *F* values obtained by analysis of variance for the remaining 11 MMPI Scales.<sup>2</sup>

Inspection of Table 1 reveals that the mean scale scores are all within normal limits with most subgroups peaking on Scales 4 (*Pd*) or 9 (*Ma*) which would be typical for adolescent samples (Hathaway & Meehl, 1951). Table 2 shows that despite the general similarity among groups indicated by Table 1 there are also marked differences. Males obtained significantly higher scores on Scales 2 (*D*), 5 (*Mf*), 7 (*Pt*), and 9 (*Ma*) than females,

<sup>2</sup> Analysis of variance tables have been deposited with the American Documentation Institute. Order Document No. 7396 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$2.50 for microfilm or \$1.75 for photocopies, and make checks payable to Chief, Photoduplication Service, Library of Congress.



whereas females scored higher on 0 (*Si*). Negroes got significantly higher scores than whites on Scales *L*, *F*, 2, 5, 8, and 9. On Scale 1, males were significantly higher than females ( $t = 1.99, p < .05$ ) and Negroes were significantly higher than whites ( $t = 3.27, p < .01$ ). Analysis of Scale 3, on the other hand, disclosed that whites were higher than Negroes ( $t = 2.23, p < .05$ ). No differences were found on any scale with regard to the class factor.

Table 2 also indicates that there are some significant interactions. The sex by race interaction on the *L* scale reflects the fact that Negro males obtained much higher *L* scores than white males ( $t = 6.97, p < .01$ ) while Negro females obtained significantly higher *L* scores than white females ( $t = 3.22, p < .01$ ), but with considerably less difference between the means. The sex by class interaction on the *K* scale pertains to the finding that upper-class females got higher *K* scores than upper-

TABLE 1  
MMPI SCALE MEANS AND STANDARD DEVIATIONS BY SEX, RACE, AND SOCIAL CLASS

MMPI scales	Males						Females					
	White			Negro			White			Negro		
	1-2	3	4	1-2	3	4	1-2	3	4	1-2	3	4
<i>L</i>	47.03 5.75	46.87 6.28	48.03 5.58	53.63 7.94	55.87 8.62	54.80 8.06	50.90 5.72	51.03 6.33	49.03 8.34	56.27 7.62	52.73 9.15	53.17 8.19
<i>F</i>	54.17 5.09	55.87 7.89	54.33 5.61	57.40 8.68	56.57 8.87	57.63 7.77	54.10 6.67	54.33 6.76	54.10 6.25	56.13 7.69	58.27 8.86	57.03 6.88
<i>K</i>	52.27 7.61	51.90 8.37	49.90 8.46	51.40 7.68	56.03 7.86	54.33 8.12	55.20 7.34	54.10 8.68	52.87 8.27	56.77 7.92	51.20 8.99	52.23 7.84
<i>Hs</i>	51.37 8.52	51.23 7.44	51.50 10.01	55.10 8.29	57.10 8.63	57.03 10.29	52.00 6.26	51.23 4.65	51.63 8.99	53.60 7.57	51.60 9.35	52.40 8.57
<i>D</i>	54.90 11.09	53.20 10.86	51.23 9.92	58.00 13.19	57.07 10.29	55.37 12.42	49.23 10.38	50.20 8.21	50.77 7.16	52.33 8.45	52.53 12.89	53.50 9.58
<i>Hy</i>	55.27 6.80	56.57 6.91	56.53 10.29	55.13 8.69	56.77 8.75	54.60 7.44	57.40 7.97	56.60 8.22	55.37 6.79	54.10 6.54	52.57 12.29	52.53 9.32
<i>Pd</i>	60.97 10.85	59.53 10.67	59.77 8.84	59.80 10.73	62.40 8.21	60.03 8.98	60.40 9.51	60.80 9.91	59.70 9.33	59.00 8.93	55.90 5.23	58.97 7.46
<i>Mf</i>	54.60 7.77	56.43 9.18	53.20 10.71	57.27 7.58	55.03 7.46	55.23 8.54	46.37 7.28	46.00 8.46	50.83 10.76	52.33 7.28	51.80 9.71	53.60 6.38
<i>Pa</i>	54.00 8.03	54.80 10.67	52.33 7.89	53.00 9.44	52.90 9.60	53.30 7.41	55.30 11.02	57.63 8.41	55.50 10.73	50.63 10.03	57.43 10.77	54.70 8.10
<i>Pl</i>	57.77 9.91	58.20 10.65	56.67 9.48	59.67 11.86	56.07 8.78	59.70 7.35	56.97 8.53	56.40 8.27	57.73 9.49	54.33 7.12	56.00 8.75	54.93 8.30
<i>Sc</i>	58.60 11.19	61.07 12.09	57.00 10.06	62.00 10.32	61.17 10.78	64.03 7.03	56.97 9.42	57.80 10.99	59.50 9.43	59.60 9.35	62.77 12.20	60.13 10.21
<i>Ma</i>	60.10 11.48	61.53 11.97	59.23 11.81	65.30 11.65	62.13 9.74	66.83 8.24	56.70 11.48	60.30 9.82	58.69 10.78	59.40 9.58	62.37 10.55	62.33 8.08
<i>Si</i>	48.47 10.01	49.83 8.55	49.57 7.74	49.90 6.38	49.27 5.29	49.10 7.62	49.30 6.38	49.27 8.64	53.63 10.02	51.00 8.50	52.57 8.44	52.67 6.27

TABLE 2

F VALUES FROM ANALYSES OF VARIANCE OF MMPI SCALES BY RACE, SEX, AND SOCIAL CLASS

Source	L	F	K	D	Pd	Mf	Pa	Pl	Sc	Ma	Si
Sex (A)	2.10	0.17	1.56	9.80**	1.51	32.71**	3.25	4.17*	1.14	4.67*	5.75*
Race (B)	49.75**	11.61**	1.19	8.09**	0.72	10.94**	1.59	0.32	7.55**	10.95**	0.59
Class (C)	0.26	0.35	1.11	0.22	0.05	0.34	2.08	0.16	0.55	0.69	1.09
A × B	5.52*	0.13	3.40	0.19	2.07	4.34*	0.39	2.21	0.13	0.40	0.63
B × C	0.96	0.08	0.30	0.01	0.13	0.56	0.72	0.20	0.11	1.40	0.65
A × C	0.06	0.09	3.41*	1.34	0.47	3.06*	1.50	0.35	0.15	1.13	1.10
A × B × C	2.16	1.05	1.27	0.05	1.49	1.09	0.61	1.36	2.07	0.42	0.70

\*  $p < .05$ .\*\*  $p < .01$ .

class males ( $t = 2.90$ ,  $p < .01$ ), whereas there was no difference between middle- and lower-class males and females ( $t = .82$  and  $0.28$ , respectively). Two significant interactions were found on Scale 5: sex by race which is a consequence of Negro females obtaining higher scores than white females ( $t = 3.94$ ,  $p < .01$ ), but Negro and white males obtaining approximately equal scores ( $t = 0.12$ ); and sex by class which reflects the findings that upper-class males' scores did not differ significantly from those of lower-class males ( $t = 1.03$ ), while lower-class females tended to get higher scores than upper-class females ( $t = 1.80$ ,  $.10 > p > .05$ ). These apparently contrasting results for the Scale 5 sex by class interaction are actually an artifact associated with the reversed scoring by sex of this scale. In other words, the high scores for upper-class males and the low scores for upper-class females both would be considered as indicating more identification with feminine interest patterns than is the case with the lower-class groups.

### DISCUSSION

There is a remarkable degree of similarity between these results and previous findings with prisoner and psychiatric samples. All studies have consistently demonstrated that Negroes obtain higher scores than whites on the clinical scales of the MMPI. It is tempting to conclude, as others have (e.g., Ball, 1960), that Negro high school students are more maladjusted than white students. It is not clear, however, that this conclusion necessarily follows from the data. A basic limitation

is that relations between scale scores or configurations and behavior have not been explicated with that high a degree of precision. Furthermore, alternative explanations are possible. As noted earlier, Dahlstrom and Welsh (1960) have suggested that results of Negro-white comparisons are a function of "known effects of socioeconomic inequities." No support of this hypothesis, however, can be adduced from the results of the present experiment, as no significant differences were attributable to the class factor.

One might argue that the crudity of the classification scheme used in this study may have played some part in the lack of positive findings with respect to class or status. However, it should be noted that there were really two class controls, parental occupation and test subjects' educational level. These criteria are analogous to those employed by Nelson (1952) with adults, namely, occupation and education. One might also point to the generally higher clinical scale scores for Negroes and ask *à la* Dahlstrom and Welsh (1960) whether this does not suggest that this group is actually of lower status than the white subjects, despite the attempted matching. A proponent of this viewpoint would, however, face serious difficulties in explaining why the two groups did not differ on  $K$  scores, especially as  $K$  has been shown to be significantly positively correlated with status.

A more general hypothesis may account for the findings, namely, that racial MMPI differences are culturally determined. More specifically, one could state that the Negro high school seniors' higher scores on  $L$ ,  $F$ , and the various clinical scales are a consequence



## Affect:

1. Hostile client behavior will evoke hostile interviewer behavior.
2. Friendly client behavior will evoke friendly interviewer behavior.
3. Hostile client behavior will evoke interviewer anxiety.

## METHOD

*General Procedure*

Four client roles of the dominant-friendly client, the dominant-hostile client, the dependent-friendly client, and the dependent-hostile client were fabricated and student actors were trained to play these roles.<sup>4</sup> Each client-actor was presented in counter-balanced order to 34 interviewers-in-training for half-hour interviews. The interviewers did not know that they were seeing actors. They were told that their clients were on the waiting list of the University Counseling Center and that the purpose of the experiment was to test the efficacy of a 30-minute interview while at the same time reducing the waiting list. The interviewers were instructed to adopt their typical interview styles. They were told to conduct what would amount to an initial interview in which their task would be either to effect disposition of the presenting problem during the interview or to make recommendations for subsequent disposition. There was considerable evidence to indicate that the interviewers believed quite deeply that they were seeing real clients. Worried consultations with supervisors, deep concerns about the decisions on the disposition of the cases, and detailed referral instructions to the Counseling Center were the rule.

In order to prevent the interviewers from talking to each other about their experiences, three precautions were taken. First, they were told not to discuss the study with anyone, especially their fellow students. Second, their own supervisors were available to discuss the interviews in order to relieve any anxiety that might have been aroused. Last, each interviewer saw all four actor-clients within a 24-hour time span, and was told the real nature of the study immediately after his fourth interview. There were no indications that leakage had occurred.

The roles fabricated for the client-actors were designed to represent four possible combinations of control (dominance-dependence) and affect (friendliness-hostility). Each actor was instructed to present a specific critical incident to the interviewer 20 minutes after the interview had begun. Descriptions of the four roles follow:

*Dominant-friendly client.* This client appears warm, friendly, gets along with people, and likes people. He feels comfortable manipulating people, meeting people, and working with people in a commanding and leading way. He is a namedropper and is apt to use people for his own purposes. In the interview situation he takes over. He shows acceptance to some extent but always keeps control. He presents his problem in a casual, rather flippant manner acting as though it were not really a problem. He just wants his own thinking clarified. He initiates conversation, takes the lead often, talks about what he wants to talk about, and does not always respond to the interviewer's directing. He comes in with a preconceived plan of what he wants the interviewer to do. He uses the interviewer as a tool. After 20 minutes he attempts to pressure the interviewer into intervening for him by convincing others that his preconceived plans are appropriate.

*Dominant-hostile client.* This client attempts to dominate, manage, and directly lead the interview. He acts confidently, tends to be boastful, assertive, aggressive, and rejects the interviewer's statements. He frequently engages in open sarcasm. He feels that other people are clods to be manipulated, and this includes the interviewer. He views others with disdain and talks of others as being without minds of their own, following the crowd. He does not come voluntarily for help but is responding to pressure from others that he seek personal help. At 20 minutes after the beginning of the interview he attempts to get the interviewer to confirm his views and his way of life and to make a direct statement that he is not in need of help.

*Dependent-friendly client.* This client is submissive, docile, and conforming during the interview. He asks the interviewer for direction and leans on him. He wants the interviewer to think well of him and at times appears to be overconventional. He expects great things from the interviewer as he does not like to make decisions by himself. No matter what the interviewer does, he believes that the interviewer will help him. He acts nervous and unsure of himself; he smiles throughout the interview and is willing and agreeable. He wants to talk things over and benefit by other's (the interviewer's) experience. He wants his life so structured that he will know ahead of time what to expect in all situations. Twenty minutes after the interview has begun, he suggests that time is running out and asks for the interviewer's evaluation and help. He persists in attempting to force the interviewer to be specific in his advice.

*Dependent-hostile client.* This is a client who is suspicious, distrustful, bitter, complaining, yet guilty, and self-effacing. He submits to the interviewer but is resentful of him. He has considerable hostility and exhibits it in a veiled way. He needs people, needs their reinforcement, but drives them away and sees his lack of interpersonal success as a function of others. He is uncomfortable in the interview situation. He gives information when pressed by the interviewer but doesn't take the lead himself. After

<sup>4</sup> The principal author is indebted to Peter D. Russell (1961) who first demonstrated to him the efficacy of using student actors as clients. The second author was stimulated to investigate this technique by Willis E. Dugan (1959) who uses actor-clients as a training method.

20 minutes, he presses the interviewer for advice but then rejects any constructive remarks of the interviewer.

### *Training the Actors*

Through the cooperation of the director of theater at the University of North Dakota, a group of student actors was obtained for audition. These actors were presented with specific role descriptions including interview behavior, feelings to be expressed to the interviewer, presenting problems, personal histories and family background material. The auditioning procedure required each actor to play a specific client role with one of the co-investigators acting as interviewer. From the group of actors auditioned, four were chosen to play the specific roles.

Six training sessions were held with each actor in which his role was discussed. After each discussion, the actor again played the role with one of the co-investigators acting as interviewer. During these sessions, the investigators tried to anticipate the kinds of interviewer behavior the actor might meet and prepare him for as many contingencies as possible. During each role playing session, the second investigator acted as observer. When the actors were judged to have met the specific role definitions set up by the investigators, they were pretested by having them engage in interviews with interviewers not used in the study. Their behavior during these sessions was rated by six observers. An actor was considered to be successfully prepared if five of the six observers indicated that he had met the role definitions.

### *Interviewers*

Thirty-four interviewers-in-training were used as subjects in the present study. All were graduate students in psychology at the University of North Dakota, ranging in age from 21 to 50. Their experience levels were commensurate with what one might expect in a graduate program in psychology, that is, ranging from a one-semester clinical practicum to several years of field experience.

### *Dependent Variables*

**Observer ratings.** Each interview was observed through a one-way mirror by a judge trained to rate affect, control, and anxiety as demonstrated by the interviewer. The instrument used to rate interviewer affect and control was a modified form of the Interpersonal Check List (ICL) published by Leary (1956). The ICL was modified in two ways: (a) The referents of all items were made interview specific. For example, the item "Encourages others" was changed to read "Was encouraging to the client"; the item "Spoils people with kindness" was changed to read "Was too kind to the client for the good of the interview." (b) Rather than simply checking each item to indicate the presence or absence of the specified behavior, the judge's response was changed

to a three-point rating scale to allow for increased discrimination.

Interview anxiety was measured by a behavioral anxiety check list developed by the investigators. The instrument required the observer to make a judgment every 2 minutes about interviewer anxiety as manifested in the six following behavioral categories: disturbances in mechanics of delivery, disturbances in content of speech, excessive silence, excessive smiling or laughing, extraneous movement, and involuntary reflex disturbances.

The observers were four advanced graduate students in clinical and counseling psychology who did not participate in the study as interviewers. They were not aware of the hypotheses being tested but did know that the clients were in fact accomplices. Since it was not practical to have more than one observer per interview, each observer was randomly assigned to 34 interviews. In order to promote inter-observer reliability, the observers were trained by having them rate trial interviews simultaneously. Training was considered completed when an inter-observer mean correlation of .88 on the ICL was obtained for the same interview which occurred after rating eight trial interviews. Reliability was easier to achieve for the ICL than for the behavioral anxiety check list. On the latter instrument it was found that training produced intraobserver consistency but did not increase interobserver agreement. The ordering of anxiety scores produced by the observers on the last three trial interviews was identical despite the fact that there were some large disparities between the absolute number of anxiety items checked by each judge. Training was discontinued at this point and it was decided that the anxiety hypothesis would be tested by transforming observer ratings on this instrument into standard scores, so that each rating could be considered as a deviation from the mean of the observer who had produced it.

## RESULTS

### *Actor-Client Consistency*

While the actors may have reached the criterion performance before the study was initiated, the question could be raised as to whether they were successful in maintaining their specified roles when exposed to a

TABLE 1  
MEDIAN RATINGS OF ACTOR-CLIENT  
ROLE CONSISTENCY

Client role	Dominance	Friendliness
Dominant-hostile	4.0	1.0
Dominant-friendly	4.0	4.0
Dependent-hostile	2.0	3.0
Dependent-friendly	1.0	4.0



TABLE 2  
MEANS OF OBSERVER RATINGS OF  
INTERVIEWER BEHAVIOR

Client role	ICL <i>M</i> Dom- inance	ICL <i>M</i> Friend- liness	<i>M</i> Anxiety
Dominant-hostile	11.39	6.10	.09
Dominant-friendly	12.75	18.03	-.01
Dependent-hostile	14.81	16.64	.03
Dependent-friendly	15.59	24.79	.06

variety of interviewer behaviors. In order to answer this question, a set of master tapes was developed which consisted of 48 randomly selected 5-minute interview excerpts (12 excerpts per actor; 1 each day for the length of study) in random order. The master tapes were presented to 12 judges who were not otherwise connected with the study and who did not know that the clients were in fact experimental accomplices. The task of the judges was to rate the behavior of the client on two five-point scales representing control and affect. Table 1 presents the median dominance and friendliness scores assigned to the actors. It can be seen that all actors were judged to have been exhibiting their specified role behavior except for the "dependent-hostile client." In this case while the median judgments indicated submissiveness, neither hostility nor friendliness was consistently evident.

#### Observer Ratings of the Interviewers

The means of the observer ratings of interviewer behavior are presented in Table 2.

TABLE 3  
ANALYSIS OF VARIANCE OF ICL RATINGS  
OF INTERVIEWER DOMINANCE

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Client control	1	333.80	8.275*
Client affect	1	38.97	.827
Subjects	33	97.01	
Control $\times$ affect	1	3.36	.111
Control $\times$ subjects	33	40.34	
Affect $\times$ subjects	33	47.14	
Affect $\times$ control $\times$ subjects	33	30.27	
Total	135		

\*  $p < .005$ ; one-tailed test.

TABLE 4  
ANALYSIS OF VARIANCE OF ICL RATINGS  
OF INTERVIEWER FRIENDLINESS

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Client control	1	2545.69	14.90**
Client affect	1	3430.52	15.97***
Subjects	33	4740.36	
Control $\times$ affect	1	120.39	.942
Control $\times$ subjects	33	170.80	
Affect $\times$ subjects	33	214.77	
Affect $\times$ control $\times$ subjects	33	127.77	
Total	135		

\*\*  $p < .001$ ; two-tailed test.

\*\*\*  $p < .0005$ ; one-tailed test.

For each dimension of the ICL a separate treatment  $\times$  treatment  $\times$  subjects design (Lindquist, 1953) was performed. From Table 3 it can be seen that the hypotheses concerning control were supported. Dominant client behavior was judged to have evoked significantly more interviewer dependence than was the case for dependent clients, regardless of condition of affect.

The hypotheses concerning interviewer affect were also supported. From Table 4 it can be seen that client friendliness was judged to have evoked significantly more interviewer friendliness than was the case for client hostility. However, unpredicted significant effects on this dimension were also found for client control; that is, client dominance was judged to have evoked less friendliness on the part of the interviewer. An examination of the means reveals that most of the effect of client control on inter-

TABLE 5  
ANALYSIS OF VARIANCE OF RATINGS  
OF INTERVIEWER ANXIETY

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Client control	1	.0002	.0002
Client affect	1	.052	.066
Subjects	33	1.410	
Control $\times$ affect	1	.142	.130
Control $\times$ subjects	33	.776	
Affect $\times$ subjects	33	.790	
Affect $\times$ control $\times$ subjects	33	1.095	
Total	135		

viewer affect was mediated by the dominant-hostile client. For as was demonstrated in Table 1, the dependent-hostile client was not very hostile and therefore evoked what could be considered an elevated friendliness score from his interviewers. With regard to interviewer anxiety, it was hypothesized that client hostility would evoke significantly more interviewer anxiety. Table 5 reveals that this hypothesis was not confirmed.

### DISCUSSION

In general, the results of this study support the position that psychotherapy should be viewed as a "reciprocally contingent interaction" (Jones & Thibaut, 1958). In such an interaction, both participants are variable responders, each alert to the incoming cues from the other and each in turn acting as a partial cause of the other's behavior. Others (Kanfer, 1961; Krasner, 1958) have noted that therapists influence the productions of their clients even though attempts at influence may not be consciously initiated. The present study suggests that clients may evoke reciprocal behaviors from their therapists even though this influence may not be perceived. It is further suggested that, unlike countertransference, which is idiosyncratic to the therapist's personality structure, evoked behavior should also be considered as a function of the real stimulus qualities of the therapeutic interaction.

With reference to anxiety, the results of this study do not support the work of Russell (1961) who found that a "hostile" actor-client produced greater interviewer anxiety than did a "friendly" actor-client. However, two important differences between the studies should be noted which would lend greater weight to Russell's results: Russell used a greater number of anxiety indices than were used in the present study; and, Russell's actor-client stimulus was unconfounded by dominance and dependence. All clients in the present study, including those designated as friendly produced management problems for the interviewer. Hence, the anxiety they evoked may have been uniformly high, preventing discrimination by the observers.

Two cautions affect the nature of the gen-

eralizations that can be made from this study. The first of these is that implied by the reliance on observer ratings for the dependent variables. A question might be raised about the specific effect of the treatment conditions upon the observers. That is, the investigators recognized the possibility that the observers themselves were being influenced by the stimulus characteristics of the actor-clients to such an extent that their own evoked feelings may have colored their perceptions of the interviewers' behaviors. Had this been the case, the findings would continue to have some meaning in terms of the hypotheses. However, the *N* of the sample would have to be reduced from 34 (interviewers) to 4 (observers).

While the interviewers used in this study did vary in experience from those with one semester's clinical practicum to those with several years of field experience, the question can be raised as to whether more experienced counselors and therapists would have responded in a similar way. It is the subjective impression of the investigators that more accomplished interviewers would experience the same personal reactions although they would be better able to control their actual interview behavior. In spite of this caution, however, the results should be useful in the preparation of therapists and counselors-in-training for the purposes of identification, understanding, and management of their own evoked behavior in the interview situation.

### REFERENCES

- BANDURA, A., LIPSHER, D., & MILLER, P. Psychotherapists' approach-avoidance reactions to patients' expressions of hostility. *J. consult. Psychol.*, 1960, 24, 1-8.
- BOLLES, R. C. The usefulness of the drive concept. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.
- CUTLER, R. Countertransference effects in psychotherapy. *J. consult. Psychol.*, 1958, 22, 349-356.
- DUGAN, W. E. Coached client: A technique for supervised counseling practicum. Minneapolis: University of Minnesota, Counselor Education Program, 1959. (Mimeo)
- FOA, J. G. Convergences in the analysis of structure of interpersonal behavior. *Psychol. Rev.*, 1961, 68, 341-353.



- JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958.
- KANFER, F. H. Comments on learning in psychotherapy. *Psychol. Rep.*, 1961, 9, 681-699.
- KENNELLY, T. E. Editorial comment on psychological trends. *J. counsel. Psychol.*, 1961, 8, 194.
- KOCH, S. Psychological science versus the science-humanism antinomy: Intimations of a significant science of man. *Amer. Psychologist*, 1961, 16, 629-639.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 55, 148-170.
- LEARY, T. *Multi-level measurement of interpersonal behavior*. Berkeley, Calif.: Psychological Consultation Service, 1956.
- LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald Press, 1957.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MASLING, J. M. The influence of situational and interpersonal variables in projective testing. *Psychol. Bull.*, 1960, 57, 65-85.
- ORNE, M. T. The nature of hypnosis: Artifact and essence. *J. abnorm. soc. Psychol.*, 1959, 58, 277-299.
- ROSENTHAL, R. On the social psychology of the psychological experiment: With particular reference to experimenter bias. Paper read at American Psychological Association, New York, 1961.
- RUSSELL, P. D. Counselor anxiety in relation to clinical experience and hostile or friendly clients. Unpublished doctoral dissertation, Pennsylvania State University, 1961.
- STRUPP, H. H. *Psychotherapists in action*. New York: Grune & Stratton, 1960.

(Received June 6, 1962)

## LANGUAGE PREDICTABILITY AS A FUNCTION OF PSYCHOTHERAPEUTIC INTERACTION<sup>1</sup>

STANLEY FELDSTEIN AND JOSEPH JAFFE

*Research Department, William Alanson White Institute, New York City*

An inverse relationship was hypothesized between number of psychotherapeutic interviews and therapists' errors in predicting their patients' speech. Recorded speech samples were obtained from Ss before their 1st psychotherapeutic interviews, and immediately after their 10th, 20th and 30th interviews. The samples were transcribed verbatim, mutilated according to the Cloze Procedure and sent to the Ss' therapists and to 2 groups of therapists who served as controls. Instructions were to guess the deleted words. The returned transcriptions were scored for proportions of incorrect guesses. A trend analysis of the data did not support the hypothesis, although the therapists were found to be better predictors of their own patients' speech than of other patients' speech.

The redundancy of a language refers to the difference between the amount of information that could be conveyed by most efficient use of the language and that amount conveyed by actual use. This concept, derived from information theory, has been found increasingly useful in tackling some of the problems of language structure and linguistic communication. Its possible usefulness in exploring the process of psychotherapy is suggested by the fact that the latter may be thought of as an interplay of similar communicative systems (Jaffe, 1958).

One technique, proposed by Shannon (1948, 1951), for obtaining a rough estimate of redundancy requires that a native speaker correctly guess a linguistic event—for example, a letter or word—on the basis of a preceding sequence of such events. In other words, the technique uses successive guesses of a single speaker. A modification of this technique, called the Cloze Procedure (Taylor, 1953), uses a single guess from each of a group of speakers.

In a discussion of information theory, Miller (1954) pointed out that an implication of the guessing procedure is that the more

an individual knows about the source and content of the text being guessed, the greater will be his success, and, hence, the greater will be the estimate of redundancy. On this basis, Miller predicted that the longer a therapist interacted with a patient, the greater would be the redundancy estimates from his attempts to guess the patient's speech. The present experiment tested the following restatement of Miller's hypothesis: There is an inverse relationship between number of psychotherapeutic interviews and the proportion of incorrect guesses in therapists' attempts to predict their patients' speech.

### METHOD

The subjects were male and female patients from an outpatient psychotherapy clinic who were seen by the experimenter on four different occasions during the course of their treatment. The first occasion ( $O_1$ ) occurred before their first psychotherapeutic interview, the second occasion ( $O_2$ ) occurred between their tenth and eleventh interviews, the third ( $O_3$ ) between their twentieth and twenty-first interviews, and the fourth ( $O_4$ ) between their thirtieth and thirty-first interviews. The strongest test of the hypothesis would have been to have seen the same subjects on each of these four occasions. Unfortunately, this procedure was not feasible. Of the 22 subjects who participated in the fourth occasion, 4 were seen on all four occasions, 14 were seen on three occasions, 3 on two occasions, and 1 on one occasion. To allow for the use of a repeated measurement analysis, each of the 18 subjects who had missed one or more occasions was matched, on the basis of a vocabulary test score, sex, and age, with another subject who had participated in the missed occasion or occasions. Thus, the data for each occasion were based on the

<sup>1</sup> The study was supported by Grants M-4548 and M-4571 from the National Institute of Mental Health. The authors are indebted to Jacob Cohen for aid in the statistical analysis, to Diane Farrell for her processing of the data, and to Doris Slawson for her unfailing cooperation in the arranging of appointments. The authors are also grateful for the participation of the candidates in training at the clinic of the William Alanson White Institute.



participation of 22 subjects, of whom 7 were males and 15 were females.

All the subjects were of native birth. The mean ages of the groups on the four occasions ranged from 23.6 to 24.5, with standard deviations that ranged from 4.4 to 5.1. On a brief measure of verbal intelligence (Thorndike & Gallup, 1944) the mean scores of the four groups ranged from 15.1 to 15.4, with standard deviations from 2.4 to 2.7. All the means are at about the 85 percentile in terms of a "standard voting population" (Thorndike & Gallop, 1944).

Each subject was seen individually and allowed 5 minutes to tell stories about four pictures.<sup>2</sup> The stories were electronically recorded and enough of the stories were transcribed to make one typed page. Verbatim transcriptions, without editing or punctuation, were used. The transcribed portion was then mutilated according to the Cloze Procedure (Taylor, 1953), that is, every fifth word was deleted. The mutilated typescript was sent, with the subject's name on it, to the subject's therapist with instructions to guess the deleted words and return immediately. Copies of the same typescript, without the subject's name, were also sent with similar instructions to two other therapists who acted as controls. The return dates of the typescripts provide the only check on the cooperation of the therapists, and seem to indicate that, for the first occasion, the therapists complied with the instructions. Twenty-one male and three female therapists took part in the experiment and were rotated so that they were in the experimental group for the prediction of their own patients' speech and acted as controls for their prediction of other patients' speech.

Upon their return, the typescripts were scored in terms of correct and incorrect guesses. To be correct a guessed word had to be the same as the original word in the passage. Words different from the corresponding original words and uncompleted deletions were counted as incorrect guesses. In order to obtain a score independent of the size of the transcribed speech sample, the number of wrong guesses was divided by the total number of deletions in the sample. For each subject, then, there were three scores: the proportion of wrong guesses made by his therapist and by each of two controls.

## RESULTS

The hypothesis stated that, compared with those of the control groups, the proportion of incorrect guesses of the experimental group of therapists shows a consistent decrease from the first to the fourth occasions. The hypothesis was tested by a trend analysis using a complex repeated measurement design (Table 1).

It was thought unnecessary to transform

<sup>2</sup> Cards 2, 8 GF, 12 BG, and 10 from the Thematic Apperception Test were used as the stimulus material.

TABLE 1

ANALYSIS OF REPEATED MEASUREMENT OF THE PROPORTION OF INCORRECT GUESSES OF THE EXPERIMENTAL (E) AND CONTROL (C) GROUPS OF THERAPISTS OVER FOUR OCCASIONS

Source	df	MS	F
Matches (M)	21	.0288	
Occasions (O)	3	.0228	
Groups (G)	2	.0821	7.20**
Linear O $\times$ G	2	.0062	1.24
Linear E $\times$ C <sub>1,2</sub>	1	.0106	2.12
Quadratic O $\times$ G	2	.0108	2.16
Quadratic E $\times$ C <sub>1,2</sub>	1	.0278	5.56*
O $\times$ M	63	.0176	
G $\times$ M	42	.0114	
O $\times$ G $\times$ M	126	.0050	

Note.—The G  $\times$  M mean square was used as the error term for evaluating the differences among Groups. The O  $\times$  G  $\times$  M mean square was used to obtain the remaining Fs. In addition to the quadratic component, there was also a significant cubic component which cannot be interpreted.

\* $p < .05$ .

\*\* $p < .01$ .

the proportions for the analysis because there were no marked fluctuations in the denominators (total number of deletions) of the original ratios. Analysis of the linear component of the curve of the experimental group compared to that of the combined control groups (Figure 1) yielded a nonsignificant  $F$  of 2.12. Comparison of the quadratic component of the two curves did, however, yield a significant difference ( $F = 5.56$ ). Further com-

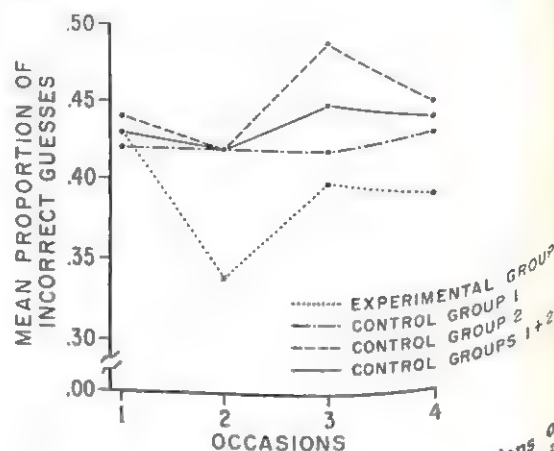


FIG. 1. The trends of the mean proportions of incorrect guesses for the therapists in the experimental and control groups across four occasions. (The means of the summed proportions of the two Control Groups were used to obtain the curve of Control Groups 1 + 2.)

parisons, using *t* tests, between the means of the first and second, second and third, and third and fourth occasions for the experimental group, indicated that its proportion of incorrect guesses decreased significantly on the second occasion and increased significantly on the third occasion.

The significant between groups *F* (7.20) was further analyzed by means of Duncan's new multiple range test, using a protection level of .99 (Edwards, 1960, pp. 138-140). The results indicated that the mean proportion of incorrect guesses of the experimental group across the four occasions was significantly lower than that of either control group even when the mean of the first occasion is included. The means of the control groups did not differ significantly from each other. Thus, although the therapists' predictions of their own patients' speech did not improve

consistently from occasion to occasion, they were, in general, significantly better than their predictions of the speech of other patients.

One question that might be raised is whether it was easier to predict the speech of same- rather than mixed-sex groups of subjects. To explore this possibility the scores for the male and female subjects within each occasion were investigated separately. Inspection of the means and standard deviations in Table 2 suggested that the predictions for the separate-sex groups were not markedly different from those of the mixed group.

#### DISCUSSION

Although the results do not support the hypothesis, they are somewhat difficult to interpret. In general, the therapists were better able to predict the speech of their own patients than that of other patients. If, as Miller suggested, accuracy of prediction is considered a measure of the predictor's knowledge of the source, then it might be said that this finding had a high antecedent probability. On the other hand, Miller also suggested that increased accuracy of prediction would imply, with reference to the psychotherapeutic situation, "better understanding" (Miller, 1954, p. 137) on the part of the predictor. It may, however, be more fruitful to conceptualize the present finding in a different way. The fact that the experimental passages were transcriptions of spontaneous, unpunctuated speech suggests that through interaction with their patients the therapists learned the vocabulary and sequential dependencies of their patients' speech well enough to be able to simulate them. Whether such simulation is a measure of understanding remains to be tested, as does the provocative question of whether quite similar speech patterns imply cognitive similarity.

Another finding of the study was that the increase in accuracy of prediction occurred at some point within the first 10 interviews. One general question raised by this finding concerns the duration of interaction required in order that the participants predict each other's speech with better than chance accuracy. Again, does the increase in accuracy

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE PROPORTION OF INCORRECT GUESSES OF THE EXPERIMENTAL (E) AND CONTROL (C) GROUPS ON EACH OCCASION (O) AND FOR THE MALE AND FEMALE SUBJECTS ON EACH OCCASION

Group	Subject sex	Statistic	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	O <sub>4</sub>
E	Mixed	<i>M</i>	.43	.34	.40	.39
		<i>SD</i>	.10	.07	.09	.11
C <sub>1</sub>	Mixed	<i>M</i>	.42	.42	.42	.43
		<i>SD</i>	.09	.15	.10	.10
C <sub>2</sub>	Mixed	<i>M</i>	.44	.42	.49	.45
		<i>SD</i>	.09	.10	.13	.13
E	Male	<i>M</i>	.46	.35	.43	.37
		<i>SD</i>	.08	.05	.09	.13
	Female	<i>M</i>	.41	.34	.39	.40
		<i>SD</i>	.10	.08	.09	.10
C <sub>1</sub>	Male	<i>M</i>	.47	.45	.48	.42
		<i>SD</i>	.10	.15	.09	.12
	Female	<i>M</i>	.40	.40	.39	.44
		<i>SD</i>	.08	.15	.09	.09
C <sub>2</sub>	Male	<i>M</i>	.41	.42	.54	.46
		<i>SD</i>	.06	.06	.06	.13
	Female	<i>M</i>	.46	.42	.47	.44
		<i>SD</i>	.10	.11	.14	.13



occur gradually or suddenly? The answers to such questions are relevant to both theories of cognition and of psychotherapeutic interaction.

The fact that the predictions of the experimental group were poorer on the third occasion than on the second occasion cannot be adequately interpreted. It may be that by the twentieth interview some kind of change had occurred within the therapeutic relationship which made the patients' speech less predictable. Perhaps a more likely conjecture is that the loss of accuracy reflected decreased motivation on the part of the therapists. Some support for the conjecture seems to be provided by the concomitant decrease in accuracy of one of the control groups, and by the fact that the therapists were not as prompt in returning the deleted word sheets (as they had been earlier in the experiment).

In brief, then, the results indicate that the relationship between number of psychotherapeutic interviews and therapists' ability to predict their patients' speech is a complex one. The results also indicate, however, that therapists do become better predictors of the speech of their own patients than of other patients. The knowledge is simple, but is important for at least two reasons. On the one hand, it has been said that what patients learn in psychotherapy is to talk their therapists' "language." The implication is that the patients of a group of therapists of the same training milieu learn to speak the "language" of that group. Were such a conjecture true, it might have been expected that the control

therapists of the present experiment would have become as able to predict the speech of the patient sample after the latter had had 10 sessions as did the experimental therapists. The results do not support the conjecture.

On the other hand, the results suggest that the Cloze Procedure is a rather simple, objective way of measuring change in the therapeutic relationship. The correlates of such change have yet to be explored. It would be useful, for example, to relate change in therapists' ability to predict their patients' speech to the length of time their patients continue treatment, or to reports of progress from both therapists and patients.

#### REFERENCES

- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1960.
- JAFFE, J. Language of the dyad: A method of interaction analysis in psychiatric interviews. *Psychiatry*, 1958, 21, 249-258.
- MILLER, G. A. Information theory and the study of speech. In B. McMillan, D. A. Grant., P. M. Fitts, & F. C. Frick, W. S. McCulloch, G. A. Miller, & H. W. Brosin (Eds.), *Current trends in information theory*. Pittsburgh, Pa.: Univer. Pittsburgh Press, 1954.
- SHANNON, C. E. A mathematical theory of communication. *Bell Sys. tech. J.*, 1948, 27, 379-423.
- SHANNON, C. E. Prediction and entropy of printed English. *Bell Sys. tech. J.*, 1951, 30, 50-64.
- THORNDIKE, R. L., & GALLOP, G. E. Verbal intelligence of the American adult. *J. gen. Psychol.*, 1944, 30, 75-85.
- TAYLOR, W. L. Cloze Procedure: A new tool for measuring readability. *Journalism Quart.*, 1953, 30, 415-433.

(Received July 9, 1962)

## RELATIVE ACCURACY OF ACTUARIAL PREDICTION, EXPERIENCED CLINICIANS, AND GRADUATE STUDENTS IN A CLINICAL JUDGMENT TASK<sup>1</sup>

LAWRENCE C. GREBSTEIN<sup>2</sup>

*Veterans Administration Hospital, Lexington, Kentucky*

3 groups of judges, varying in amounts of clinical experience, predicted Wechsler-Bellevue IQ scores from Rorschach psychograms in competition with a multiple regression equation. The results indicated that: (a) There was no significant difference between the degree of accuracy of the judgments of the equation and of the clinicians. (b) The judges and the multiple regression equation were able to judge IQ scores significantly better than chance. (c) Clinical predictive accuracy, as indicated by high positive correlations between predicted and actual IQ scores and low average error scores, did not increase significantly with amount of clinical experience.

Despite the availability of extensive evidence, the question of "clinical versus statistical prediction" is still an unsettled one in the hearts of many clinical psychologists. Although Meehl (1954) has offered an impressive array of data for the superiority of the actuarial approach to clinical judgment, some persons have resisted accepting his conclusions. For example, Holt (1958) feels that the evidence in favor of the actuarial approach may be a function of the experimental design, in which the clinician is at a disadvantage, rather than a function of the actual superiority of the statistical method. He suggests that an interesting study could be made of simultaneous attempts to predict

the same criteria from the same data by clinicians and statisticians who have gone through the same preliminary steps. In the same vein, Meehl (1956) states, in reference to a review of the studies in this area:

I have reservations about some of these studies; I do not believe that they are optimally designed to exhibit the clinician at his best; but I submit that it is high time that those who are so sure that the "right kind of study" will exhibit the clinician's prowess, should do this right kind of study and back up their claim with evidence.

One of the purposes of this study is to carry out Holt's suggestion and accept Meehl's challenge. Accordingly, the present study is concerned with the ability of clinicians to judge Wechsler-Bellevue IQ scores from Rorschach psychograms in competition with the ability of a linear multiple regression equation to perform the same task. Although this is not a function in which clinical psychologists are specifically trained, it is one which they frequently do. Also, since the experimental design includes judges with differing amounts of clinical experience, it permits an evaluation of judgment in relation to experience. It is an attempt to clarify the many previous studies in this area in two ways. First, the task is equally amenable to either statistical or clinical analysis, but favors neither approach; and second, the results can be evaluated in terms of a more objective criterion of validity.

<sup>1</sup> This paper is based upon the author's master's thesis, in partial fulfillment of the requirements for the master's degree, University of Kentucky. It was presented in summarized form at the 1961 American Psychological Association Convention in New York. The author is indebted to the University of Kentucky Computing Center Director and to John Donahoe, data, and to Richard Griffith, Chief Research Psychologist of the Veterans Administration Hospital, Lexington, Kentucky, for his many helpful suggestions in the preparation of this manuscript. The author also wishes to express his gratitude to the staff of the Psychology Service of the Veterans Administration Hospital, Lexington, for their cooperation, participation in the study, and for making their files available.

<sup>2</sup> Presently a graduate student at the University of Kentucky and a Clinical Psychology Trainee at the Veterans Administration Hospital, Lexington, Kentucky.



## HYPOTHESES

The following hypotheses are being tested in the present study:

1. There will be no significant difference between the degree of accuracy of the judgments of the linear multiple regression equation and the sophisticated group of clinical psychologists.

2. Both the judgments of the sophisticated group of clinicians and the linear multiple regression equation will be significantly better than chance.

3. Clinical predictive accuracy, as indicated by high positive correlations between predicted and actual IQ scores, will increase with amount of clinical experience.

## METHOD

### Judges

Three groups of judges, five in each group, were used in this study. The *naïve* group consisted of graduate students who had completed an introductory course in Rorschach administration and scoring. They had not been trained in advanced Rorschach interpretation and had no formal clinical experience with the Rorschach. The *semisophisticated* group of judges was composed of graduate students who had completed both an introductory and interpretation course in the Rorschach and who were in the second-year level of training in the Psychology Service at the Lexington Veterans Administration Hospital. The *sophisticated* judges were professional clinical psychologists with the PhD degree and a minimum of 5 years clinical experience. All judges were trained in the Beck scoring system for the Rorschach Test.

### Psychograms

Approximately 100 IQ scores of patients tested in 1954 and 1955 were taken from the files of the Lexington Veterans Administration Hospital, and the distribution and range of these scores were determined. From this original population, 30 IQ scores, ranging from 57 to 142, with the same approximate distribution as the original sample, were selected for the judges' task. The mean of this sample was 91.56 and the standard deviation was 20.34. Each record was screened to insure that the following criteria were met:

1. The Wechsler-Bellevue Intelligence Scale and Rorschach must have been administered to the patient within 1 week of each other.

2. In no case were any of the judges taking part in the study to have administered or scored any of the tests or collaborated in the final report. In addition, the scoring was checked for each of the

records and any irregularities in the psychograms were corrected. In a few cases, some of the tests originally selected had to be replaced because one or more of the above criteria had not been satisfied.

The psychograms were typed, put in random order, coded by number, and presented to each judge. The judges were instructed to estimate each IQ. Each judge was allowed as much time as he wanted. The judges received the psychograms successively, so that none of them were doing the task at the same time, and each judge was asked not to consult anyone else in making his judgments.

### Multiple Regression Equation

Three hundred and nineteen additional IQ scores were chosen from the same population of patient files as that used for the judges' task; the additional scores were then put in the form of a frequency distribution with class intervals of 10 IQ points. From these, 100 were selected in proportion to the original distribution. The range, 47 to 131, was approximately the same as for the sample used by the judges. As in the previous sample, each psychogram was checked for errors. A multiple regression equation was derived, based on 10 Rorschach factors from the Beck scoring system. These were:  $R$ ,  $W$ ,  $M$ ,  $F+$ ,  $Z$ ,  $A$ ,  $P$ , range of content ( $RC$ ), and range of determinants ( $RD$ ). From this equation, the most significant variables were used to determine the best linear equation; a more detailed description of this procedure appears in the Results section. Finally, scattergrams were plotted for each of the 10 Rorschach factors, to determine if their relationship to intelligence was linear or curvilinear. This was done to justify the use of a linear multiple regression equation as the appropriate one for prediction.

## RESULTS

### Comparison of Judgments with Multiple Regression Equation

The best linear equation to be used to compete with the judges was determined by successively dropping out factors from the 10-variable equation until the residual sum of squares for regression  $N$  (group of variables dropped) represented a significant loss when compared with regression  $N$ . This follows the procedure suggested by Quenouille (1952), who states:

To test whether any particular variable contributes significantly in the prediction of a dependent variable, it is necessary first to carry out an analysis of variance when the variable is included and then to repeat the analysis when the variable is missed out. In this manner, the sum of squares due to the inclusion of the variable may be estimated and tested, using the variance-ratio test (p. 75).

TABLE 1  
PREDICTIVE ACCURACY OF THE JUDGES AND THE MULTIPLE REGRESSION EQUATION

	Naive	Semi-sophisticated	Sophisticated	Multiple regression equation
Correlation	.50*	.65**	.68**	.56**
Average error	17.77	13.92	14.49	13.53

\*  $p < .05$ .  
\*\*  $p < .01$ .

This procedure also allows for the "education" of the equation, so that a more meaningful comparison can be made between its performance with that of the judges. The analysis indicated that there was no significant loss when the 10-factor equation was reduced to 5 factors or when the  $W\%$  factor was eliminated from the 5-factor equation.<sup>3</sup> However, there was a significant loss when  $RC$  (range of content) was dropped from the 5-factor equation. Thus, the following equation was chosen to compete with the judges. Significant  $t$ 's (.05 and .01 levels) were associated with all of the factors in the equation.

$$IQ = 57.54 + 3.73 M + .20 F + \% + .25Z + 1.20 RC$$

The predictions made by this equation had a .56 correlation with actual IQ and an average error of 13.53.

The group mean correlations were calculated by correlating the actual IQ scores with the mean predicted IQ (based on five judges' predictions for each group) for each psychogram. It is important to note the method for these computations, since a careful inspection of Table 2 might suggest that the group means are too high when compared with the correlations of the individual judges in each group.

The average error score for each individual

<sup>a</sup> Two 1-page tables, containing an analysis of variance for significance of regression (Figure A) and final multiple regression equation (Figure B) have been deposited with the American Documentation Institute. Order Document No. 7454 from ADI Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

judge (Table 3) represents the mean deviation from the 30 actual IQ scores of his 30 predictions of IQ. The group mean average errors were computed in similar fashion to the group mean correlations. That is, the difference was found between actual IQ and the mean predicted IQ for all five judges within each group for each psychogram.

Table 1 summarizes a comparison of the performance of the equation with the judges.

The results summarized in Table 1 indicate that Hypotheses 1 and 2 were supported. Both the sophisticated group of judges and the multiple regression equation were able to predict Wechsler-Bellevue IQ scores significantly better than chance, and there was no significant difference between the judgments of the sophisticated group of judges and the multiple regression equation.

### Effect of Experience on Judgments

Table 2 shows the individual and mean correlations of predicted with actual IQ for the three groups of judges.

TABLE 2  
CORRELATION OF PREDICTED IQ WITH ACTUAL IQ  
FOR INDIVIDUAL JUDGES AND GROUPS

	Naive	Semi-sophisticated	Sophisticated
Mean	.50*	.65**	.68**
Judge 1	-.02	.51**	.56**
Judge 2	.34	.54**	.57**
Judge 3	.37	.57**	.57**
Judge 4	.53**	.61**	.58**
Judge 5	.56**	.64**	.68**

\*  $p < .05$ .  
\*\*  $p < .01$ .



TABLE 3  
AVERAGE ERROR SCORES FOR INDIVIDUAL  
JUDGES AND GROUPS

	Naive	Semi- sophisti- cated	Sophisti- cated
Mean	17.77	13.92	14.49
Judge 1	22.17	14.93	14.56
Judge 2	16.87	14.70	16.40
Judge 3	14.90	13.36	12.97
Judge 4	16.67	12.10	14.86
Judge 5	18.87	14.50	13.63

Hypothesis 3 was tested by an analysis of variance (Table 4) based on predicted IQ scores for the three groups of judges. The analysis indicated that there was no significant difference among judges between groups. Therefore, it was concluded that the ability to judge IQ from Rorschach psychograms does not improve with additional clinical experience. The only significant source of variance was between Rorschach psychograms.

Fisher's  $z$  transformation test (Edwards, 1954) for the significance of the difference among correlations indicated that the mean correlation for the three groups of judges did not differ significantly from one another. Since the analysis of the results indicated that there were no statistically significant differences between the groups, it was necessary to reject the hypothesis that accuracy of clinical

TABLE 4  
ANALYSIS OF VARIANCE FOR DIFFERENT  
LEVELS OF JUDGES

Source of variation	df	MS	F
Between judges (J)	14		
Groups (G)	2	901.04	1.55
Between judges same groups	12	580.27	
Within judges	435		
Rorschachs (R)	29	2,409.31	23.27**
G $\times$ R	58	88.06	.85
Pooled J $\times$ R	348	103.54	
Total	449		

\*\* $p \leq .01$ .

judgment increases with amount of clinical experience.

*Cues used by judges.* Additional analyses of the data indicated that the interjudge reliability was much higher for the more experienced judges (semisophisticated and sophisticated groups) than it was for the naive judges. That is, experienced judges tended to resemble each other in their performance, making similar errors. The range of interjudge correlations for the three groups was: Naive,  $-.07$  to  $+.72$ ; Semisophisticated,  $+.55$  to  $+.85$ ; and Sophisticated,  $+.54$  to  $+.80$ . Individual multiple regression equations derived for each judge showed that the more experienced judges frequently used the same four variables, ( $M$ ,  $Z$ ,  $F+\%$ , and  $RC$ ) as the competitive equation. The multiple regression coefficients associated with all the judges ranged from  $+.84$  to  $+.98$ , indicating that the judgments of the individual judges could be predicted with a high degree of accuracy.

An inspection of the individual predictions of all the judges and the multiple regression equation suggested that predictive accuracy was consistently lowest for the higher IQ scores. The average error computed between the five highest IQ scores and their corresponding predictions by all the judges and the equation was 23.48, as compared with their mean average error of 14.92 for the entire task sample (including the five highest IQ scores). It appears that the relationship between IQ scores and Rorschach variables breaks down considerably within the upper range of IQ scores, and this tends to lower the overall correlations of the judgment process for both the clinical judges and the multiple regression equation.

## DISCUSSION

Any investigation of clinical judgment must face certain unavoidable difficulties. For example, we might ask about this study: Is predicting IQ scores from Rorschach psychograms a true indication of clinical judgment? It is very possible that this task, while allowing for a more rigorous experimental design and statistical analysis, is so limited that the results do not yield any information

about the complex process of clinical judgment. It is difficult to determine if a clinician is doing the same thing when he makes a judgment on the basis of assorted information as he is when he is predicting one test score from another test, as the judges did in this study.

The question might be raised as to whether many of the judges were taking the task seriously enough and whether, with more time and involvement, they might have "beaten" the equation. The following statement was inserted in the instructions, in order to motivate them: "Since you will be competing with other clinical judges and also with a multiple regression equation, try to be as accurate as you can." While no objective measure of the judges' motivation is available, most of them were very eager to know how well they did on the task. Most of the judges kept the psychograms several days, but it is difficult to estimate how much of this time was actually spent doing the task. Several of the judges reported that they referred to the psychograms on different occasions, "mulling it over" in their minds. They reported, after this delay, actually taking from 30 to 60 minutes writing down their estimates. Not only did the judges know that the psychograms to be rated were those of patients, but many of the judges worked at the hospital from which the psychograms were chosen. Although every effort was made to limit extraneous information being given to the judges, it is possible that they may have a tendency to lower their total IQ estimates on the basis of their familiarity with the hospital population. This factor was controlled as a possible source of error by deliberately using a wide but representative range of IQ scores. In order to determine whether or not the judges' predictions were biased by their familiarity with the hospital population, correlations were computed between the mean predictions for each group for all the actual IQ's except the five highest. The correlations increased for the two more experienced groups, but not for the naive group, when the highest IQ's were removed from the sample. However, since the largest increase occurred in the multiple regression equation, it cannot be concluded that bias was responsible for the judges' predictions.

It might be argued that depriving the judges of the original protocols represents a departure from the manner in which a clinician who does not depend upon actuarial methods arrives at his judgment of intelligence. This procedure was deliberately used, and the rationale behind it is based on some findings reported by Todd (1954). On a similar task, involving the estimating of IQ from the Rorschach test, Todd found that, while the IQ ratings using total protocols were significantly more highly correlated with IQ than psychogram ratings alone for the group, only one of the 10 judges significantly increased his ratings. He concluded that the addition of verbal cues may increase some clinicians' ability to judge intelligence from the Rorschach, but there is no evidence that this is generally the case. He also found that 39 sophomore college students were able to rate 78 Rorschach verbal protocols (without the scoring records) according to the degree of verbal complexity believed to be reflected in the protocol. The correlation of the median ratings with IQ was  $+ .58$ , as compared with  $+ .47$  for the clinicians' ratings based on formal factors alone. This suggested that verbal factors were at least as valuable as formal factors as cues to intelligence. On the basis of these findings, the present investigator felt that allowing the judges to use the complete protocol would probably not increase their accuracy significantly and it might tend to confound the results. That is, one would not be able to say that ability to do the task was dependent upon the amount of experience or training in clinical psychology. In addition, there was no good way to include the qualitative differences of the verbal protocols into the equation.

Taft (1955) has indicated that judgmental ability is positively correlated with intelligence. Therefore, the objection might be raised that, in the present study, the success of the judges in predicting IQ scores is as much a function of their intelligence as it is of their relative amounts of experience. In order to see if this was the case, a rank order correlation coefficient was computed between the mean general aptitude score (consisting of quantitative and verbal percentiles) of the Graduate Record Examination and accuracy



of prediction for the 10 graduate students of the naive and semisophisticated groups of judges. The rank order correlation coefficient ( $Rho$ ) obtained was  $-.28$ . Thus, it seems that, for this study, the judges' ability to predict IQ's was not a function of their intelligence alone.

Finally, Hammond (1955), in his discussion of the clinical method, has included data from the study by Todd (1954), which is comparable to some of results obtained in the present study. For example, Todd found that the median correlation coefficient between Rorschach-estimated IQ and Wechsler-Bellevue IQ for 10 experienced clinicians was  $+.47$ . Although Hammond states that this correlation is "better than chance and fairly impressive," it is considerably lower than the  $+.68$  correlation coefficient obtained by the the experienced clinicians in the present study. In terms of the relationship of various Rorschach factors to IQ, Todd reports the following correlations:  $+.31$  with  $R$ ,  $+.33$  with  $M$ ,  $-.30$  with  $F\%$ , and  $+.28$  with  $F+\%$ . The present study indicated that  $M$  and  $F+\%$  correlated positively and significantly with IQ, suggesting some stability of the relationship between these Rorschach factors and Wechsler-Bellevue IQ. As in the present study, Todd predicted the responses of the clinical judges by developing separate multiple regression equations (based on the four factors best for each judge) for each of the 10 clinicians. The median was  $+.85$  and the individual correlations ranged from  $+.74$  to  $+.92$ . In the present study, the corresponding range of correlation coefficients was from  $+.84$  to  $+.98$  for all judges, regardless of their experience. These data support Hammond's (1955) contention that "... the multiple correlation model which predicts

that the clinician combines the data from the Rorschach in a linear, additive fashion is a good one . . ." (261).

#### FINAL REMARKS

It is obvious, from the previous discussion, that the results of this study must be interpreted with reservations. At the same time, they must not be viewed either pessimistically or capriciously. These findings have been obtained.

1. Experienced clinical psychologists can make accurate estimates of intelligence from Rorschach summaries, but there is room for improvement.

2. A linear multiple regression equation can do the same task with approximately equal but no greater success.

3. It appears that clinical experience beyond an introductory course in Rorschach does not significantly increase judgmental or predictive ability.

#### REFERENCES

- EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.
- HAMMOND, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.*, 1955, **51**, 255-262.
- HOLT, R. Clinical vs. statistical prediction: A re-formulation and some new data. *J. abnorm. soc. Psychol.*, 1958, **56**, 1-12.
- MEEHL, P. *Clinical vs. statistical prediction*. Minneapolis: Univer. Minnesota Press, 1954.
- MEEHL, P. *When shall we use our heads instead of the formula?* Paper read at American Psychological Association, Chicago, August 1956.
- QUENOILLE, M. *Associated measurements*. New York: Academic Press, 1952.
- TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1-23.
- TODD, F. S. *A methodological analysis of clinical judgment*. Unpublished doctoral dissertation, University of Colorado, 1954.

(Received July 12, 1962)

## PURDUE PEGBOARD AS A PREDICTOR OF THE PRESENCE AND LATERALITY OF CEREBRAL LESIONS<sup>1</sup>

LOUIS D. COSTA, HERBERT G. VAUGHAN, JR., ERIC LEVITA, AND NORMAN FARBER

*Albert Einstein College of Medicine, Yeshiva University*

A portion of the Purdue Pegboard Test taking 3 minutes to administer was given to 80 patients on a neurology service of a general hospital. With optimum cutoff scores it was possible to indicate the presence and laterality of brain damage in 70% of the cases. Brain damage without regard to lateralization was correctly determined in 90% of cases with 7.5% false negatives and 2.5% false positives. The test was then cross-validated on 65 consecutive admissions with small loss in predictive efficiency. The data support the use of sensorimotor rather than perceptual cognitive tests in screening for brain damage because they are less dependent upon educational background and because of the lateralizing significance of sensorimotor dysfunction.

Attempts of psychologists to diagnose the presence and laterality of cerebral lesions have been criticized for failure to use tests which can be quantified reliably, to control for age and education, and to use adequate validating and cross-validating procedures (Meyer, 1961; Yates, 1954). The usefulness of psychodiagnosis of brain damage must also be considered in terms of cost in time and money, the ease with which diagnosis can be made without such psychological study, and the base rate incidence of neurological disorder in the population sampled.

Costa and Vaughan (1962) and Vaughan and Costa (1962) in a survey of perceptual cognitive and sensorimotor performance in brain damaged patients found deficits in performance on all tests in brain damaged as compared to control groups. The perceptual cognitive tests, however, correlated highly with education. In a group of patients heterogeneously distributed with regard to education, the predictive efficiency of such tests would be lowered unless separate norms were set up for groups differing in amount of education.

An approach which overcomes effects of educational level involves intraindividual comparisons on a large variety of tests. Such techniques reveal specific or general tests

defects correlated with brain damage (Reitan, 1962). The results are obtained at the cost of the time involved in a complete evaluation and thus such batteries are not feasible as screening procedures.

An alternative approach to the psychological assessment of brain damage lies in the use of sensorimotor tests (Semmes, Weinstein, Ghent, & Teuber, 1960; Teuber & Weinstein, 1954; Yates, 1954). These tests offer measures of function essentially independent of educational achievement. We found the Purdue Pegboard (Science Research Associates, 1948) to be a test which is correlated with other indicators of sensorimotor performance (Vaughan & Costa, 1962) and independent of educational level in normals. We decided, therefore, to assess its use as a screening device in detecting the presence and laterality of cerebral lesions.

The addition of a short screening test of high accuracy to the routine battery of psychological tests given in many clinical settings would be of great value. The presence of positive findings would suggest medical evaluation or re-evaluation, perhaps even in instances where brain damage was not suspected. Additional psychological techniques also relevant to brain damage but not often used routinely might then also be used. The small cost of administering a short sensorimotor test routinely would be amply repaid by the detection of even a few unsuspected cases of brain damage.

<sup>1</sup>This project was supported by NIH Grants 2M6418 C5 and B 3356.

A version of this paper was presented at the American Psychological Association Meeting at St. Louis, September 1962.



TABLE 1  
CORRELATION OF AGE AND EDUCATION WITH INTELLECTUAL AND SENSORIMOTOR TESTS IN A  
GROUP OF 26 NON-BRAIN-DAMAGED PATIENTS

Tests	Age	Education	Education with age held constant
Age	—	—	—
Mill Hill vocabulary	—	-.56	—
WAIS block design	-.43	.69	.60
Pressure threshold (right hand)	-.48	.60	.56
Two-point discrimination (right hand)	.74	-.61	-.40
Finger oscillation (right hand)	.22	-.30	-.23
Purdue Pegboard (right hand)	-.57	.39	.14
	-.39	.30	.11

### PROCEDURE AND SUBJECTS

The peg placement part of the Purdue Pegboard test was administered for the left hand, right hand, both hands simultaneously, exactly as prescribed in the examiner's manual (Science Research Associates, 1948). The actual testing time was 90 seconds, 30 seconds for each condition. With required demonstrations and instructions the entire testing procedure required approximately 3 minutes. The number of pegs placed correctly was recorded for the left and right hands, and the number of pairs placed correctly was recorded for the simultaneous condition.

Two samples of patients were tested. The validating sample consisted of 80 patients referred to the neuropsychology laboratory for psychodiagnostic evaluation. They were independently diagnosed by clinical neurological examination, electroencephalography, and neuroradiographic procedures. All patients considered testable were capable of reaching the testing room on foot or in a wheel chair. Psychological test performance on 54 of these patients has previously been reported (Costa & Vaughan, 1962;

Vaughan & Costa, 1962). The cross-validation sample consisted of 65 consecutive testable admissions to the neurology service. Their diagnoses were determined by the same procedure as in the validating sample. The brain damaged patients exhibited a wide variety of acute and chronic disorders commonly seen on an adult neurology service. Cases of neoplasm, traumatic injury, degenerative, vascular, and infectious disease were all represented. The control groups contained patients admitted to the neurology service with lesions either in the peripheral nervous system or below the level of the thoracic spine. Some patients were diagnosed as having non-neurologic lesions. Ten patients could not be categorized on the basis of neurological evaluation and were excluded from consideration. Table 2 displays the sex, age, and education data of the samples studied.

### RESULTS AND DISCUSSION

On the basis of inspection of the frequency distributions of the scores of patients in the

TABLE 2  
SEX, AGE, AND EDUCATION OF VALIDATION AND CROSS-VALIDATION SAMPLES

Group	Sex			Age		Education	
	N	M	F	M	SD	M	SD
Validation sample							
Control	26	10	16	51.92	16.73	8.53	2.90
Left lesion	20	11	9	57.50	13.80	8.05	4.44
Right lesion	20	10	10	57.75	8.51	7.65	3.07
Bilateral lesions	14	8	6	47.14	12.93	8.17	3.68
Cross-validation sample							
Control	15	9	6	48.33	14.29	9.60	3.84
Left lesion	10	7	3	47.10	15.35	9.10	2.84
Right lesion	10	7	3	55.10	14.20	6.10	2.88
Bilateral lesions	20	17	3	51.60	15.79	7.80	3.38

validating sample alone the following cutoff rules<sup>2</sup> were derived:

Subject below age 60

1. Call brain damaged if one or more of the following apply: left < 11, right < 13, simultaneous < 10, left > right, right > left + 3. If none apply call non-brain-damaged.

2. Call lesion (a) left if left > right and (b) right if right > left + 3. If called brain damaged but a and b do not apply, call brain damaged bilateral.

Subject age 60 and above

1. Call brain damaged if one or more of the following apply: left < 10, right < 10, simultaneous < 8, left > right, right > left + 3. If none apply call non-brain-damaged.

2. Call lesion (a) left if left > right and (b) right if right > left + 3. If called brain damaged but a and b do not apply, call brain damaged bilateral.

The cutoff rules were constructed to provide maximal predictive accuracy with no special consideration given to minimizing false posi-

<sup>2</sup> The proportion of left-handed patients was too small to permit the development of separate norms.

tive or false negative results. Obviously other cutoff rules could be determined to minimize either kind of error at the cost of increasing the other.

*Predictive Efficiency*

The cutoff rules were then applied to the independent cross-validation sample of consecutive admissions.

Table 3 shows that the predictive accuracy of the test with regard to the presence of brain lesions is approximately 90% in both the validation and the cross-validation samples. In all 135 cases there were only six false positives and eight false negatives. We assume that the cross-validation sample of consecutive admissions provides an estimate of the base rate incidence of cases falling into each diagnostic category. Seventy-three percent of all cases are brain damaged so a predictive accuracy of 73% can be attained by calling every case brain damaged. The accuracy of 89% obtained with the Purdue

TABLE 3  
ACCURACY OF PREDICTION OF PRESENCE OR ABSENCE OF BRAIN LESIONS  
WITH THE PURDUE PEGBOARD

Sample	Neurological diagnosis	N	Prediction		Percent accuracy
			Correct	Incorrect	
Validation	Control	26	24	2	92
	Left lesion	20	16	4	80
	Right lesion	20	19	1	95
	Bilateral lesion	14	13	1	93
	(All brain damaged)	(54)	(48)	(6)	(89)
	Total	80	72	8	90
Cross-validation	Control	15	11	4	73
	Left lesion	10	10	0	100
	Right lesion	10	9	1	90
	Bilateral lesion	20	19	1	95
	(All brain damaged)	(40)	(38)	(2)	(95)
	Total	55 <sup>a</sup>	49	6	89
Validation cross-validation combined	Control	41	35	6	85
	Left lesion	30	26	4	87
	Right lesion	30	28	2	93
	Bilateral lesion	34	32	2	95
	(All brain damaged)	(94)	(86)	(8)	(92)
	Total	135 <sup>a</sup>	121	14	90

<sup>a</sup> Ten nondiagnosed cases omitted.



TABLE 4

ACCURACY OF PREDICTION OF LATERALITY OF BRAIN LESIONS WITH PURDUE PEGBOARD

Sample	Neurological diagnosis	N	Prediction				Accuracy		Percent accuracy
			Non-lesion	Left lesion	Right lesion	Bilateral lesion	Correct	Incorrect	
Validation	Control	26	24	1	—	1	24	2	92
	Left lesion	20	4	11	1	4	11	9	55
	Right lesion	20	1	—	13	6	13	7	65
	Bilateral lesion	14	1	2	3	8	8	6	57
	(All brain damaged)	(54)	(6)	(13)	(17)	(18)	(32)	(22)	(59)
	Total	80	30	14	17	19	56	24	70
Cross-validation	Control	15	11	1	1	2	11	4	73
	Left lesion	10	0	8	—	2	8	2	80
	Right lesion	10	1	—	6	3	6	4	60
	Bilateral lesion	20	1	4	7	8	8	12	40
	(All brain damaged)	(40)	(2)	(12)	(13)	(13)	(22)	(18)	(55)
	Total	55	13	13	14	15	33	22	60
Validation cross-validation combined	Control	41	35	2	1	3	35	6	85
	Left lesion	30	4	19	1	6	19	11	63
	Right lesion	30	2	—	19	9	19	11	63
	Bilateral lesion	34	2	6	10	16	16	18	48
	(All brain damaged)	(94)	(8)	(25)	(30)	(31)	(54)	(40)	(57)
	Total	135	43	27	31	34	89	46	66

Pegboard is significantly higher\* ( $p < .05$ ) than that obtained calling all cases brain damaged.

The predictive accuracy with regard to lateralization is shown in Table 4. Seventy percent accuracy in the validation sample shrinks to 60% accuracy in the cross-validation sample. If cases were assigned to categories of lateralization on the basis of their proportional representation in the cross-validation sample of consecutive admissions an accuracy of lateralization of only 27% would result. The accuracy of lateralization using the Purdue Pegboard is significantly higher ( $p < .01$ , see footnote 3) than 27%.

Thus both with regard to the presence and laterality of lesions the Purdue Pegboard offers a significant improvement over prediction from base rate.

### Errors in Prediction

Analysis of the false positive cases with regard to the presence of cerebral lesions suggests the test is weakest in differentiating patients with conversion symptoms and with peripheral nerve or cord disease affecting the

\*  $t$  tests for the significance of differences between percentages.

upper extremities from brain damaged patients. No generalization seemed to apply to the false negative cases. Many of the bilateral lesion patients who were misdiagnosed as having lateralized lesions did, in fact, give independent clinical evidence of greater impairment in the hemisphere predicted but nevertheless were considered as errors with regard to lateralization. The test does not discriminate about 25% of patients with lateralized lesions from patients with bilateral lesions. No explanation of this fact was apparent from a consideration of the cases involved. It is to be noted that only 1 of 60 patients with lateralized lesions was actually mislateralized.

### Implications

Even though the Pegboard test has been demonstrated to be useful as a screening device, its predictive usefulness in other settings remains to be determined. Situations where the base rate incidence of diagnostic groups is different, such as in outpatient clinics, state hospitals, or other institutions, would change predictive efficiency and might require new cutoff scores.

No consideration was given here to the usefulness of the test in diagnosing "difficult"

as compared to "easy" cases. While some of the patients tested were hemiparetics whose lesions could be lateralized almost at sight, other correctly lateralized patients showed no obvious sensorimotor disturbance.

No claim is made that the Purdue Pegboard test is successful in differentiating non-brain-damaged psychiatric patients from patients with cerebral lesions. This, except for a few cases of conversion hysteria where the test did poorly, was not a relevant diagnostic question in the setting where the test was validated.

The Purdue Pegboard test is not a substitute for a large battery of tests assessing various aspects of dysfunction. It, in itself, does not offer promise for correctly making the large variety of inferences that can be made from such batteries (Reitan, 1962).

In our experience, however, the Pegboard test shows promise as a screening device. It strongly suggests the superiority of tests which maximize sensorimotor and minimize verbal or intellectual operations, as single indicators of brain damage. Such sensorimotor tests have two obvious advantages: their relative independence of educational level and their ability to elicit bimanual samples of per-

formance which may indicate laterality of lesion.

#### REFERENCES

- COSTA, L. D., & VAUGHAN, H. G., JR. Performance of patients with lateralized cerebral lesions. I: Verbal and perceptual tests. *J. nerv. ment. Dis.*, 1962, 134, 162-168.
- MEYER, V. Psychological effects of brain damage. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961. Pp. 529-565.
- REITAN, R. M. Localization of brain lesions and the problem of simultaneously covarying neurological variables in human patients. Paper read at Eastern Psychological Association, Atlantic City, April 1962.
- SCIENCE RESEARCH ASSOCIATES. SRA examiner's manual for the Purdue Pegboard. Chicago, Ill.: SRA, 1948.
- SEMMES, JOSEPHINE, WEINSTEIN, S., GREENT, LILA, & TEUBER, H. L. *Somatosensory changes after penetrating brain wounds in man*. Cambridge: Harvard Univ. Press, 1960.
- TEUBER, H. L., & WEINSTEIN, S. Performance on a form board task after penetrating brain injury. *J. Psychol.*, 1954, 38, 177-190.
- VAUGHAN, H. G., JR., & COSTA, L. D. Performance of patients with lateralized cerebral lesions. II: Sensory and motor tests. *J. nerv. ment. Dis.*, 1962, 134, 237-244.
- YATES, A. J. Validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, 51, 359-379.

(Received July 12, 1962)



## EMPATHY, NEED TO CHANGE, AND IMPROVEMENT WITH PSYCHOTHERAPY<sup>1</sup>

ROSALIND DYMOND CARTWRIGHT

AND

BARBARA LERNER

*University of Illinois College of Medicine, Chicago*

*Billings Hospital, University of Chicago*

To study the influence of: the patient's need to change, the therapist's experience level, the empathic understanding of the therapist, the sex of the patient and therapist, and the amount of psychological distance between them, on psychotherapy improvement, 28 patients in client-centered counseling were tested before and after therapy with scales built from the Kelly Role Construct Repertory Test. Patients' initial need to change was found to be directly related to improvement. Therapists' final level of understanding was also directly related. Patient's need and therapist's empathy considered jointly produced a prediction model for therapy length and success. 2 success groups were found: same-sex patients of experienced therapists whose distance from him the therapist initially reduced, and opposite-sex patients of inexperienced therapists whose distance from him the therapist initially increased.

There is a growing body of evidence indicating that improvement in psychoneurotic patients takes place concurrent with psychotherapy, but as yet little consensus concerning the factors responsible for such improvement. To the majority of writers in this field, improvement is a function of specific patient variables, therapist variables, and the interactions among them. However, different writers select different factors as accounting for the major portion of the variance in therapeutic outcome. Since the research evidence is most often unreplicated, or even contradictory, and since there are still large unexplored areas, one can pick and choose among the scraps and fashion a patchwork quilt after one's own heart or theoretic commitment.

Although the various schools of psychotherapy have different views as to what constitutes the necessary and sufficient conditions of the therapeutic process, the studies which compare the respective percentages of successfully treated cases have shown rather similar results for all theoretic approaches. This frequent finding suggests that some common elements exist among the various approaches to psychotherapy and, that they might very well be more significantly related

to improvement than the elements on which the schools differ.

The evidence for two such common elements has been accruing slowly. The studies of Heine (1950) and Fiedler (1950a, 1950b) strongly suggest that one important therapist variable, independent of technique differences, is the therapist's ability to empathically understand his patient. A patient variable which is suggested by the work of Butler and Haigh (1954), Rosenthal and Frank (1956), Cartwright and Cartwright (1958), and Kirtner and Cartwright (1958), as one important to any form of verbal therapy, is the patient's initial recognized need to change. These two promising variables, *the therapist's empathy* and *the patient's need to change*, were selected to form the basic dimensions of this study. However, since we now feel that therapy depends not only on the qualities of the two participants but also on how these affect their relationship, this study attempted to tease out some of the interaction effects of these two major variables.

Part I of the study was undertaken with three primary hypotheses:

1. The degree of need to change on the part of the patient is directly related to improvement with psychotherapy.

2. The empathic understanding of the patient by the therapist is directly related to the degree of improvement in the patient with psychotherapy.

<sup>1</sup> The authors are pleased to acknowledge the very considerable help they received in clarifying the material in this paper from the critical reading of the manuscript by Ernest Haggard and Fred L. Strodtbeck.

3. These two variables considered jointly will give a better prediction of outcome than either taken singly.

After analyzing the data relative to these hypotheses which had been formulated prior to the data collection, other questions arose concerning the interaction of additional variables with those selected for this study and with the dependent variable, improvement. Part II of this paper considers the sex of the patient in relation to the sex of the therapist, the "psychological distance" the therapist puts between himself and his patient, and the experience level of the therapist. These additional variables were explored to help clarify some of the interpersonal processes underlying the results obtained from testing the major hypotheses. This study, then, involves the consideration of five variables, some of their interactions, and their relation to improvement with psychotherapy. These five variables can be organized according to a threefold scheme: (a) patient characteristics independent of the particular therapist (recognized need to change), (b) therapist characteristics independent of particular patient (therapist's level of experience), and (c) characteristics dependent upon the particular patient-therapist pair (the therapist's empathy or ability to understand the patient in his own terms, whether or not the patient and therapist are of the same sex, and the therapist's distancing of the patient).

## METHOD

### *Sample*

The subjects were 28 patients who voluntarily sought treatment at the Counseling Center of the University of Chicago.<sup>2</sup> Fourteen of the subjects were male, 14 female. They ranged in age from 19 to 43 with a mean age of 27.7. Although these patients were not formally diagnosed, according to their therapists' ratings on "severity of illness" they ranged from "very mild" to "near psychotic." Sixteen client-centered therapists were involved, and the case length varied from 6 to 116 interviews with a mean of 40.

<sup>2</sup> No selectivity of patients was exercised. Each new applicant was requested to participate in the study until the quota of 30 subjects was reached. Two patients had not completed treatment at the time the data analysis was begun.

## *Instruments*

The two instruments in this study are referred to as Scale A and Scale B. Scale A provided the data for the improvement criterion and Scale B provided the measure of the patient's need to change, the measure of the therapist's empathic understanding of his patient, and the therapist's distancing of the patient.

Scale A was administered to the therapists on two occasions, after the second interview, referred to as the pretherapy rating, and after the last interview, for a posttherapy rating. The Improvement score was the sum of four components: three change scores between the pre- and posttherapy ratings of: (a) the patient's integration; (b) his kind of organization (defensive versus open); and (c) his present life adjustment. The fourth component of the Improvement score was the therapists' final rating of the outcome of therapy. All ratings were made on nine-point scales.

Scale B differed in content for each subject. Each subject in a sense supplied his own items. This was done in an effort to insure that the scales would be highly personally relevant to each patient. The items were the personal constructs obtained by first administering Kelly's (1955) Role Construct Repertory Test to each subject. From these personal constructs which each patient introduced as constituting the important similarities and differences among the real people in his life, the first 10 discrete ones were selected. Scale B was then made up of the 10 items supplied by each patient arranged as five-point rating scales. On the first testing occasion following interview Number 2, the subject made various ratings of these items: (a) to describe himself as he is at present, and (b) to describe himself as he wants to be when therapy has been completed. The sum of the squared discrepancies between these two sets of ratings was used as the measure of his felt need to change on the 10 items of particular significance to him. The patient repeated this task at posttherapy to again describe himself as he was then.

Scale B was also the basis for the measurement of the therapist's empathic understanding of his patient. After the second interview the therapist was given a Scale B form containing the 10 items chosen by his patient. He then attempted to describe "the patient as he sees himself." The empathy measure was the squared discrepancy between the patient's self-description and the therapist's attempt to predict the patient's self-description. The therapist repeated this task at posttherapy time. This procedure was used to measure the therapist's empathic understanding after increased contact with the patient.

## RESULTS

### *Part I*

*Hypothesis 1: The degree of need to change on the part of the patient is directly related*



TABLE 1

MEAN SCORES ON PATIENTS' NEED TO CHANGE AND THERAPISTS' EMPATHY  
FOR THE IMPROVED AND UNIMPROVED PATIENTS

Variables	Improved <sup>a</sup>	Unimproved <sup>b</sup>	<i>t</i>	<i>df</i>
Need to change	47.26	20.00	3.17**	26
Pre-T empathy (E <sub>1</sub> )	15.93	16.69	—	
Post-T empathy (E <sub>2</sub> )	9.40	14.46	2.24*	26

Note.—Since the Empathy Score is based on a  $D^2$  measure, the lower the score the greater the empathy.

<sup>a</sup>  $N = 15$ .

<sup>b</sup>  $N = 13$ .

\*  $p > .02$ .

\*\*  $p > .01$ .

to improvement. Table 1 shows that Hypothesis 1 is strongly supported.

The mean need to change score was much higher for the improved than for the unimproved group.

*Hypothesis 2: The empathic understanding of the patient by the therapist is directly related to the degree of improvement in the patient with psychotherapy.* Table 1 shows that there was no significant difference between the improved and unimproved cases on their therapists' ability to understand their pretherapy self-image. However, at the close of therapy, the therapists understood the self-image of the improved patients significantly better than they did those who were unimproved. Also there was a significant gain in the therapists' empathy score between the first (E<sub>1</sub>) and second (E<sub>2</sub>) occasions—but only for the improved cases. This result showed that there was a significant relation between improvement and increased understanding. For those who were rated unimproved through therapy, the therapists made no significant gain in their understanding of them.

Perhaps the explanation of the relation between high posttherapy empathy and improvement is not the obvious one, that the therapist coming to understand his patient's own view of himself contributes to more and better therapeutic work being done. Perhaps, instead of the therapist coming to understand the patient, the patient is adopting his therapist's view of him. This alternate view might well account for the results. Such a change would make it easier for the therapist to predict how the patient "sees himself" and

also possibly be valued as a growth in insight making it more likely that the case be rated "improved." We have, then, to test: do improved cases change so that their posttherapy self-description resembles the description the therapist made of them at pretherapy time more closely than their pretherapy self-description did. The chi square test here was significant in the opposite direction. At posttherapy, 11 of the improved cases descriptions of themselves were *less* like the therapists' pretherapy descriptions of them than they had originally been, and only 4 were more like them. For the unimproved group the reverse relationship was found, 9 cases were more like their therapists' pretherapy conceptions of them and only 4 were less like them. Improvement, then, goes with a patient change away from the therapist's early conception of him. Is the therapist's posttherapy conception of the patient then less like his pretherapy self-description and more like the patient's self-conception? The answer, given by a highly significant chi square, is yes for patients rated improved. The therapists of improved cases have changed their conceptions of the patients in ways that bring them closer to the patients' own view.

Perhaps the improved cases stayed longer and so these therapists had more contact time within which to improve their understanding. Actually, the improved cases had fewer interviews ( $M = 37.33$ ) than the unimproved cases ( $M = 43.69$ , a nonsignificant difference). Thus neither a change to resemble the therapists earlier views nor the amount of contact alone accounts for the relation found here between improvement in therapy and

increase in the level of understanding of the patient by the close of therapy.

*Hypothesis 3: The two variables considered jointly will give a better prediction of outcome than either taken singly.* By taking the two variables together, improvement and length of treatment, some interesting differences were revealed. Dividing the cases at the median by length into Long (25 or more interviews) and Short (less than 25), and into Improved (Improvement score 10 or greater) and Unimproved (Improvement score 9 or less) four subgroups were formed. Table 2 gives the means for these groups on the two measures.

Schematically, then, the relations between the variables were as follows:

Length improvement	Need to change	Empathy ( $E_1$ )
Short improved	High	High
Long improved	High	Low
Short unimproved	Low	Low
Long unimproved	Low	High

In terms of the joint consideration of the patient's need to change and the therapist's empathic understanding, several generalizations can now be offered.

Therapy is short when either of two conditions obtains. (a) Both the patient's need to change and the therapist's understanding of the patient are high. Although the mean number of interviews is small, these cases leave therapy with high improvement scores. (b) The patient's need to change and the

therapist's understanding of him are both low. These cases leave therapy in an equally short time but as unimproved.

Therapy is long, on the other hand, under two other conditions. (a) The patient's need to change is high but the therapist initially misperceives him. The high degree of patient motivation to change seems to be sufficient to keep him in contact long enough for the therapist to correct his misperception and come to see the patient in his own terms ( $E_1 = 20.00$ ,  $E_2 = 8.57$ ). These are long term cases but do eventually leave improved. (b) The therapist understands the patient's self-conception but the patient doesn't really want to change. The high understanding probably has sufficient reward value to keep the patient in contact but without the internal pressure to change or recognition that change is possible, he eventually leaves therapy unimproved.

## Part II

Having reached this stage in the analysis, the study might well have been concluded. However, further questions kept occurring for which some clarification might be reached by searching the available data in new ways.

The finding that therapists varied in their level of understanding of their patients' self-conceptions after an exposure of only two interviews was not surprising. Nor was it surprising to find that some therapists, whose first attempts at understanding their patients were pretty wide of the mark, increased their

TABLE 2  
PATIENTS' NEED TO CHANGE, THERAPISTS' EMPATHY FOR FOUR IMPROVEMENT SUBGROUPS

PATIENTS' NEED TO CHANGE, THERAPISTS' EMPATHY FOR FOUR							
Group	Patients' Need to change		Therapists' $E_1$		Therapists' $E_2$		Empathy change $E_1-E_2$
	$M$	Rank	$M$	Rank	$M$	Rank	$M$
Short Improved $N = 8$	56.62	1	12.37	1	10.12	2	2.2
Long Improved $N = 7$	36.57	2	20.00	4	8.57	1	11.4
Short Unimproved $N = 6$	17.66	4	19.83	3	18.16	4	1.6
Long Unimproved $N = 7$	22.00	3	14.00	2	11.28	3	2.7



TABLE 3  
SEX OF THERAPIST-PATIENT PAIRS AND EMPATHY, IMPROVEMENT, AND DISTANCING

Variables	Same Sex <sup>a</sup>	Opposite Sex <sup>b</sup>	<i>t</i>	<i>df</i>
E <sub>1</sub>	20.21	12.35	2.278**	26
E <sub>2</sub>	12.00	11.50	—	
E <sub>1,2</sub>	8.21	.85	1.777*	26
Improvement	10.00	9.35	—	
Distancing	+ 5.21	- 6.42	2.211**	26

<sup>a</sup> *N* = 14.

<sup>b</sup> *N* = 14.

\* *p* < .10.

\*\* *p* < .05.

understanding to a high level by the time therapy had drawn to a close. However, since there was a strong relationship between the therapist's ability to understand his patient's self-conception at the close of therapy and the outcome of the case, there was interest in discovering what characteristics of the relationship made the initial understanding either easy or difficult, and what characteristics were related to an improvement in understanding through time. The first of these characteristics to be investigated was the sex of the therapy pair.

### Sex

Therapists obtained significantly higher empathy scores on the first occasion with patients of the opposite sex than with patients of the same sex. By the time therapy had been completed this difference no longer held. It seemed that the therapists had more initial difficulty understanding patients of like sex than of the opposite sex but that this handicap was overcome with time. (Therapists' empathy for clients of like sex did increase significantly  $t = 2.37$ ,  $p < .05$ ,  $df$  13.) Perhaps therapists at the beginning of their contacts with patients of the same-sex err in understanding by assuming that they are more like themselves than is warranted. This assumption of similarity would be less likely to occur with patients of the opposite sex, leaving the therapist freer from a projective set and more open to discovering how it is that the patient views himself.

### Distance

In order to test the suggested explanation of the effect of the sex pairing on empathy, a distance measure was devised. First the actual similarity in

the ratings that the therapist and patient each made to describe themselves was calculated. The squared discrepancy between these was called a measure of their "Real Similarity." Next the similarity of the therapist's description of his patient and of himself was calculated. The squared discrepancy between these two was called a measure of the therapist's "Assumed Similarity." The difference between the Real and Assumed similarity scores was used as a measure of the distance the therapist placed between himself and the patient. If he "assumed" more similarity than was "real" he was erring psychologically by bringing the patient closer to himself than was represented by the reality distance between their two self-ratings. These errors were scored with a positive sign and are referred to as errors in the direction of reducing the distance. If the therapist assumed less similarity than was real he erred in the direction of putting the patient farther from him than corresponded to the real difference in their ratings. These errors were scored with a negative sign and referred to as errors of increasing the distance.

Therapists on the first testing occasion reduced the distance with same-sex patients and increased the distance with opposite-sex patients. This tendency would seem to help to account for the finding of poorer initial empathy with same-sex than with opposite-sex patients. It appears from Table 3 that therapists potentially can understand patients of either sex equally well ( $E_2$  12.00 and  $E_2$  11.50) but that the initial assumption that the self-images of same-sex patients will be more like their own than in fact they are, temporarily delays the full operation of the empathic capacity. Although therapists also err initially with patients of the opposite sex by overemphasizing the differences between them, this negative distancing does not seem to interfere markedly with the empathy score. That is to say, therapists may incorrectly perceive the patient as very different from

themselves and still perceive the way he sees himself reasonably accurately.

### *Distance, Sex, and Improvement*

Now what is the effect of this initial distancing of same- and opposite-sex patients on their eventual therapeutic gain? Table 4 shows that the therapist's initial distancing of both same- and opposite-sex patients who were judged improved at termination was considerably more extreme than his distancing of those who were judged unimproved. Although both improved sex groups were subject to large distancing errors, these were in opposite directions. The same-sex patients who improved in therapy were initially seen by the therapist as more like himself than their own ratings would suggest. This reduction of distance seems to imply an immediate emotional acceptance of these people. In contrast, the same-sex patients who were subsequently rated as unimproved were held off emotionally at the beginning of therapy and seen by the therapist as more different from him than their own ratings placed them as being. For the opposite-sex patient the relationships were reversed. The opposite-sex patients who improved in therapy were early seen by their therapist as being very different from him—more different than their self-ratings showed them to be. This viewing the opposite-sex patient as very different seemed to imply a classification of, and insistence upon, the sex role distinction. The cross-sex patients who failed to improve were not seen by the therapist as occupying a pattern distinct from his own. They were seen as being about as much

like the therapist as their self-ratings showed them to be.

The really big difference was between same- and opposite-sex patients who improved. Same-sex patients who improved were early accepted by the therapist as very like him and opposite-sex patients who improved were early seen as very different.

### *Experience of the Therapist*

The effect of the experience level of the therapist on distancing, empathy, and improvement was investigated next. Eight of the therapists were classified as more experienced, on the grounds of having handled more than five research cases prior to the present one, and eight were classified as inexperienced, having treated fewer than five. For five of the therapists the present case was the first one. The experienced therapists made distancing errors in both directions with about equal frequency. They were as likely to err in the direction of seeing more similarity than actually present as they were to see less. The inexperienced therapists, though, had a significant bias in the direction of the negative errors. Inexperienced therapists tend to see patients as less like them than the patients' own views show them to be. Perhaps what one learns as a result of experience in doing therapy is that nothing human is really foreign to us. Or perhaps it is that the new therapist is more open to threat and more easily made anxious by seeing patients as similar to himself and so does more defensive distancing.

Now to relate these various findings to

TABLE 4  
EFFECT OF DISTANCING OF SAME- AND OPPOSITE-SEX PATIENTS ON THE  
IMPROVED AND UNIMPROVED CASES

Sex pairing	Distancing				
	Improved	Unimproved	<i>t</i>	<i>df</i>	<i>p</i>
Same sex	+14.00	-6.50	2.594	12	.02
	<i>N</i> = 10	<i>N</i> = 4	2.893	12	.02
Opposite sex	-11.85	-1.00			
	<i>N</i> = 6	<i>N</i> = 8			
<i>t</i>	4.649	.797			
<i>df</i>	14	10			
<i>p</i>	.001	<i>ns</i>			



improvement with psychotherapy, it appears that improvement was rated high for two groups of patients: same-sex patients treated by experienced therapists, and opposite-sex patients treated by inexperienced therapists. In both groups the therapists' posttherapy level of empathy ( $E_2$ ) was high and the original distancing of the patient by the therapist extreme. The experienced therapist achieved high empathy and improvement with same-sex patients with whom the psychological distance was immediately reduced and the inexperienced therapist achieved high empathy and improvement with opposite-sex patients with whom distance was immediately increased.

#### DISCUSSION

The study shows that psychoneurotic patients have the best chance for a successful treatment experience with nondirective therapy, if they come to it with a high need to change, and meet a therapist who can accurately understand the way they see themselves. Some patients meet with therapists who have difficulty perceiving them in their own terms. If they have sufficiently high motivation to change they continue in therapy, and the therapist does come to understand more accurately how it is they see themselves.

For the therapist to improve his understanding in this way he must have more access to material relevant to the task. Presumably he misperceives originally because of barriers to this material, either in himself which distort his perceptions, or because of barriers in the patient which prevent him from revealing himself accurately. Barriers of both kinds are, in all likelihood, related to defenses against the potential threat involved in conscious recognition. If threat is reduced in the patient he would be more able to communicate personally relevant material to the therapist. This in turn would enable the therapist to base his view of his patient's self-image on a deeper understanding of him. If threat is reduced in the therapist he would be freer to perceive the patient's communications without distortions.

It would seem that the inexperienced therapist is the more open to threat in this situation than is the experienced therapist, and,

in truth, it is he who tends to increase the distance with his patients. This in effect denies some of the similarity between himself and his patients which can be interpreted as a defensive maneuver to reduce his threat. This distancing might well be experienced by his patients as a message to the effect that "you are really very different from me." Such a message would probably raise the threat level of the same-sex patients but reduce it for opposite-sex patients. The experienced therapist who has less personal threat decreases distance between himself and his same-sex patients and conveys by this, "you are really much more like me than you think you are." This message from a prestigious person of the same sex probably reduces the patient's threat level. This tentative explanation would account for the high improvement and final level of empathic understanding of the same-sex patients treated by experienced therapists and of the opposite-sex patients treated by inexperienced therapists.

#### REFERENCES

- BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling in *Psychotherapy and personality change*. In Carl R. Rogers & Rosalind F. Dymond (Eds.), Chicago: Univer. Chicago Press, 1954. Pp. 55-75.
- CARTWRIGHT, D. S., & CARTWRIGHT, ROSALIND D. Faith and improvement in psychotherapy. *J. consult. Psychol.*, 1958, 5, 174-177.
- FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, non-directive, and Adlerian therapy. *J. consult. Psychol.*, 1950, 14, 436-445. (a)
- FIEDLER, F. E. The concept of the ideal therapeutic relationship. *J. consult. Psychol.*, 1950, 14, 239-245. (b)
- HEINE, R. W. An investigation of the relationship between changes and the responsible factors as seen by clients following treatment by psychotherapists of the psychoanalytic, Adlerian, and non-directive schools. Unpublished doctoral dissertation, University of Chicago, 1950.
- KELLY, G. *The psychology of personal constructs*. New York: Norton, 1955.
- KIRTNER, W., & CARTWRIGHT, D. S. Success and failure in client-centered therapy as a function of initial in-therapy behavior. *J. consult. Psychol.*, 1958, 22, 329-333.
- ROSENTHAL, D., & FRANK, J. Psychotherapy and the placebo effect. *Psychol. Bull.*, 1956, 53, 294-302.

(Received July 26, 1962)

## INTRAINDIVIDUAL CONSISTENCY IN "CREATIVE" AND "MEMORY" STORIES WRITTEN FOR TAT PICTURES<sup>1</sup>

RUTH C. WYLIE

*Goucher College*

BOYD D. SISSON

*Medical College of Georgia, Augusta*

AND EARL TAULBEE

*Veterans Administration Hospital, Lincoln, Nebraska*

108 college women wrote at 2 sittings, either following the usual (creative) instructions both times, providing reliability data, or following "creative" instructions one time, drawing stories from memory the other. Memory sources were varied in 2 subconditions. Aggression and Dependency scores were used in separate tests of the clinically assumed hypothesis that it makes no difference diagnostically whether S is being creative or drawing stories from memory. When correlating scores from 2 types of instructions, card set and session number were balanced. Intraindividual consistency is low or zero under all conditions, despite reliable scoring. Aggression rs are larger than Dependency rs. The data do not justify using the TAT as scored here for individual diagnosis.

Lindzey (1961) has recently stated that the most general assumption underlying the use of projective testing is: "If an individual is presented with a stimulus situation permitting variable responses, the particular responses that he emits will reflect his characteristic response patterns and tendencies to response" (p. 146). Although he recognizes that there are many sources of variance in projective test responses, his assumption implies that projective testing can lead to reliable and valid diagnosis of individuals only to the extent that intraindividual response consistency occurs, under specified conditions of data collection, as demonstrated by suitably designed research. Empirical exploration of Lindzey's general assumption must involve a very wide variety of actual studies, of which the present is one example.

We are here concerned with intraindividual consistency in Aggression scores and in Dependency scores from TAT protocols, when the test is administered with specified variations in instructions. The particular

sets of instructions were chosen in order to explore the proposition that remembered stories are subjected to the same motivational influences as are fantasied (creative) stories. Of course TAT testers typically do not know whether a given story is one which the subject remembers from a book, movie, or other source; or whether it is a more genuinely creative product. Nevertheless, they seem to feel that this lack of knowledge is not a serious handicap, since their scoring and interpretive procedures involve no systematic attempt to identify the sources of the subjects' stories. Amazing as it may seem, an empirical test of this practically important example of the general assumption of response consistency has not heretofore been attempted.

The findings pertinent to the main hypothesis concerning "creative" and "memory" influences must be evaluated against the test-retest reliability of the TAT when it is given under the usual creative instructions commonly employed in clinical work. Accordingly, suitably instructed control groups are included in the present study. In view of the astounding scarcity of published reliability information for the TAT, these groups give data which are interesting and valuable

<sup>1</sup> The final draft of this paper has been improved through comments by D. T. Campbell and L. Sechrest for which we wish to express appreciation. Final responsibility for the study and the opinions expressed rests, of course, with the authors.



in themselves, as well as providing control data for the present experiment.

In operational terms, the hypotheses tested in this study are:

*Hypothesis 1.* There is a significant positive correlation between Aggression scores (or between Dependency scores) obtained from TAT protocols written by the subjects at one sitting and from protocols written by the same subjects at a second sitting approximately a week later, when both sets of stories are written under the usual (creative) instructions.

*Hypothesis 2.* There is a significant positive correlation between Aggression scores (or between Dependency scores) obtained from TAT protocols the subjects have written under the usual (creative) instructions, and from TAT protocols consisting of stories which the subjects choose from memory. This correlation will be as high as that between two sets of protocols by the same subjects done under creative instructions.

In testing Hypothesis 2, two kinds of sources for the "memory stories" were used by different groups of subjects: stories provided by the experimenter at sessions prior to those in which the subjects wrote their TAT stories, and stories which the subjects chose freely from their own personal memories of movies, books, short stories, case history materials, etc.

From the clinician's viewpoint, the more immediately relevant results come from the groups who freely chose their memory stories, because the situation more closely resembles what the subjects may do in the clinical testing situation.

On the other hand, when the subjects are provided with a restricted, controlled set of possible memory stories, this means that they have all had relatively comparable influences of immediately past learning upon their choice of memory stories, as contrasted to the situation in which each subject draws her memory stories from her own unique set of readings, movie viewings, etc. Intraindividual consistencies which might be relatively obscured by sources of variable error in the "free choice" situation might appear more strongly in the situation where opportunities for choice of memory stories are made more comparable across all subjects.

Unfortunately, even the strongest trends in our results show disappointingly weak tendencies toward intraindividual consistency in Aggression and Dependency scores of the sort employed in this experiment. No support is given those who assume that suitably reliable discriminations among individuals can be made under any of the conditions used in this study.

## METHOD

*Stimulus materials.* Twelve TAT cards were chosen and subdivided into two sets of six each, called Set A and Set B. In the absence of published norms which would enable the experimenters to choose two sets of cards comparable in their drawing power for Aggression and Dependency, the sets were formed on the basis of the judgments of the two authors who are clinicians. Because this method of equating card sets was only approximate, the design of the experiment allowed for counterbalancing any effects of inequality between the sets. Set A consisted of Cards 1, 3BM, 5, 7B, 14, and 17BM. Set B consisted of 3GF, 7GF, 13G, 13B, 17GF, and 20.

For each card three stories were obtained or specially prepared. In the judgment of the two authors who are clinicians, one story for each picture represented a strongly aggressive trend, one a strongly dependent trend, and one was neutral so far as aggression and dependency were concerned.

Two mimeographed booklets were prepared, one each for Card Sets A and B. Each booklet contained 18 stories, one to a page, with the stories pertaining to the various cards intermingled randomly throughout the booklet.

*Subjects.* Results are based on the records of 108 female beginning psychology students who were divided into 10 groups, as described in the Procedure Section and in Table 1.

*Procedure.* Groups I and II were called together in separate group meetings first (called Pre-1 Session). They were treated identically except that Group I used booklets of stories prepared for Set A, while Group II used the booklets of stories prepared for Set B. In the Pre-1 Sessions, Group I and II subjects received the following instructions:

When the experimenter tells you to do so (*not now*), you are to open your booklet to page one and read the story on that page silently while she reads it aloud to you. Concentrate as much as you can while she is reading the story, because your job is to remember the story as well as you can until our next meeting. You don't need to try to remember the wording of the story, but pay special attention to the main events the story describes, what the characters think and feel, and the outcome, or probable outcome, described. As soon as the experimenter has read the story on page one with you, she will tell you to turn to page two. She will then read this story to you while you read it silently. Concen-

TABLE 1  
DESIGN OF EXPERIMENT

Group	Session Pre-1	Session 1	Session Pre-2	Session 2
I	Read memory stories for Set A	Write memory stories for Set A	—	Write creative stories for Set B
II	Read memory stories for Set B	Write memory stories for Set B	—	Write creative stories for Set A
III	—	Write creative stories for Set A	Read memory stories for Set B	Write memory stories for Set B
IV	—	Write creative stories for Set B	Read memory stories for Set A	Write memory stories for Set A
I'	—	Write "free" memory stories for Set A	—	Write creative stories for Set B
II'	—	Write "free" memory stories for Set B	—	Write creative stories for Set A
III'	—	Write creative stories for Set A	—	Write "free" memory stories for Set B
IV'	—	Write creative stories for Set B	—	Write "free" memory stories for Set A
V	—	Write creative stories for Set A	—	Write creative stories for Set B
VI	—	Write creative stories for Set B	—	Write creative stories for Set A

Note.—The subjects wrote TAT stories under varying conditions: (a) with the usual (creative) instruction, (b) with instructions to write stories selected from a group of stories they had studied previously, (c) with instructions to write stories based on any memory source they wished to use. For full explanation see text. Sets A and B are two sets of six cards each.

trate on this story so that you can remember the same things about it as you did about the story on page one. We shall proceed in this way through the 18 stories in the book. Your job will always be to try to remember the main points of the stories just as well as you can.

After a brief pause the subjects read the stories through again, this time by themselves and in a different order from that of the first reading, spending 1 minute on each story and studying each story in any way they thought would help them to remember it. Finally, they gave the stories a third silent reading in still a different order, taking 1 minute per story. At no time were they given any information that TAT pictures would be used on their next sessions.

In the sessions called Session 1 and Session 2, which were spaced about a week apart and begun about a week after the Pre-1 Session, members of these groups were treated individually. The subject was seated at a table, given the set of cards she was to use, a written copy of the directions for writing TAT stories, and a watch with which to limit herself to 7 minutes for writing of each story. For Session 1, the directions were those typically used in administering the TAT, except that the subject was told to choose for each picture a story from the booklets she had read during the Pre-Session. She was told to improvise or modify if she could not remember clearly the

story she thought most appropriate to the picture. During Session 2, the directions for writing were the same as for Session 1, except that stress was laid on being creative and avoiding stories remembered from the mimeographed booklets or from other sources such as movies or magazines.

Members of Groups III and IV were seen individually for Session 1, and wrote creative stories for Card Sets A and B respectively. The directions and procedure were identical for those of Session 2 (creative session) of Groups I and II, except that no mention was made of avoiding plots from the mimeographed booklets, since these subjects had not had experience with the booklets. For the session called Pre-2 Session, members of Groups III and IV met in separate groups and read the stories for Set A and B, respectively, in the same manner as described for the Pre-1 Session of Groups I and II. Of course the subjects in Groups III and IV could surmise that TAT cards would be forthcoming in their next session. The treatment of Groups III and IV subjects in Session 2 was identical with that of Group I and II subjects during Session 1.

It is obvious that when motivational scores from memory stories are correlated with motivational scores from creative stories across all 48 subjects in Groups I-IV, the factors of card set and session number are comparable between the two arrays of scores correlated.



Groups I' through IV' were treated just as Groups I through IV, with the following important exception: in Groups I' through IV' no story-reading preessions were held. Instead, for the sessions in which the subjects were to produce memory stories, they were instructed to base their stories on something they remembered from real life, movies, books, case history materials, etc. The exact instructions given to these subjects were:

We have said that this is a test of creative imagination. However, we should like to add the following requirement: Try to use, as a basis for each story, some situation or "plot" which you remember from a novel, short story, play, radio or TV drama, real life case history, or news story. You may write the story pretty much as you remember it, or you may modify it if you feel that is desirable in order to make it more appropriate to the picture than the literally remembered story would be. And, if you don't remember very clearly the situation you are drawing on as a source for one of today's stories, take what you do remember and use it as a point of departure in writing the story. The point is to base the story for each picture not on sheer imagination, but on some human situation you have heard about before today. . . .

In this case, too, when motivational scores from memory stories are correlated with motivational scores from creative stories across all 36 subjects in Groups I' through IV', the factors of card set and session number are comparable between the two arrays of scores correlated.

Groups V and VI were treated like the other groups, except that they were given only creative instructions at both Sessions 1 and 2, as follows:

Your task will be to tell as dramatic a story as you can for each picture. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. You may find yourself wanting to use plots from movies, books, magazines, or radio or TV dramas. Please avoid that—*be just as creative and original as you can possibly be.*

Group V wrote stories in response to Card Set A in Session 1, and to Card Set B in Session 2, while the reverse order of card sets was used for Group VI. Thus, when motivational scores from Session 1 are correlated with motivational scores from Session 2, across the 24 subjects in Groups V-VI combined, the factors of card set and type of instruction are comparable for the two arrays of scores correlated.

Before the experiment began, the subjects had been admonished to keep all aspects of the procedure completely secret, and postexperimental interviews revealed that they had apparently cooperated in this to a very high degree throughout the experiment. Each subject was asked to write

her idea of the purpose of the experiment and no subject had guessed it correctly. Prior to the experiment the subjects had been assured their anonymity would be preserved by a system of code numbers on their stories, and after their last session they were shown that all handwritten records were typed before being scored.

## RESULTS

Each story from Groups I through IV was typed on a separate page, together with its code number, decipherable only by the first author who did not participate in the scoring. The two authors who are clinicians took 50 of these stories at random and independently applied Aron's (1949) scoring scheme for two variables, Aggression and Dependency. If both scorers left a scoring unit unscored, or both scored it for the variable under consideration, this was counted as agreement. Out of 306 total possible agreements, there were 79% agreements on Aggression and 88% on Dependency. This was not considered satisfactory, so the judges scored and discussed some stories not being used in the study, and then independently scored 50 more stories selected randomly from those in the study. On this second reliability check, there was 91% agreement on Aggression out of 368 scoring units, and 93% agreement on Dependency. It was decided that this degree of reliability would be satisfactory, so the remaining stories from Groups I through VI were divided at random between the scorers. The final data for Groups I through VI are based on the combined judgments of the two scorers on the stories they had both scored, and on the separate judgment of one or the other of the scorers on the remaining stories.

Data from Groups I' through IV' were collected after the first six groups, and a psychology graduate student scored these stories, after practicing on stories of Groups I through VI until she achieved at least as high independent agreement with each of the original scorers as they had achieved with each other. (A split-half reliability check was later made, within each session of each group, and split-half reliabilities did not vary systematically as a function of group, indicating that creative-memory correlations obtained from Groups I' through IV' may

safely be compared with those obtained from Groups I through IV.)

Since there were marked differences in story length, each story was given a score obtained by dividing the number of scoring units judged as Aggressive (or Dependent) by the number of scoring units in the story. These proportion scores were summed for each subject for her six creative stories and for her six memory stories, and the two sums were used in testing the hypotheses, with respect to the motivational variable under consideration.

Table 2 presents the findings relevant to the hypotheses.

Hypothesis 1 is confirmed by the fact that data from Groups V-VI yielded a positive correlation between Aggression scores from the two sets of stories, both of which were written under creative instructions; and a positive correlation between Dependency scores in the two sets of stories, both written under creative instructions. The correlation for Aggression is significant between the .05 and .02 levels, on a two-tailed test; but the correlation for Dependency scores reaches the .05 level of significance on a one-tailed test only. While these levels of significance suffice to confirm Hypothesis 1, the small size of the coefficients gives no support to the idea that one may reliably use sets of six TAT stories for individual diagnosis of Aggression or Dependency, even when the card sets are chosen specifically for their supposed relevance to these needs, and the stories are written under a procedure rather closely

resembling the usual clinical testing procedure.

In examining the results pertinent to Hypothesis 2, one needs to consider Groups I-IV as a combined group, and Groups I'-IV' as another combined group. This is so because the memory stories written by the subjects were drawn from restricted sources in Groups I-IV, while the subjects in Groups I'-IV' were instructed to draw freely on their personal memories when writing their memory stories. Table 2 shows that, for Groups I'-IV', the correlations between Aggression scores from creative and memory stories is significantly positive, as Hypothesis 2 predicted. This  $r$  is almost the same size as the  $r$  found in Groups V-VI between Aggression scores from the two sets of creative stories. Thus, the introduction of a different set of instructions (to draw stories from personal memories) did not seem to lead to any lower intraindividual consistency than is found in two sets of stories written under the usual creative instructions. Nevertheless, the intraindividual consistency again falls short of that needed for reliable individual diagnosis.

With respect to the Dependency scores of Groups I'-IV', the results are different. Here there is an insignificantly negative correlation between scores from memory stories and creative stories. There is no support for the hypothesis that intraindividual consistency in motivational influences occurs across two sorts of processes: choosing and writing a story from one's own memories and creating a story imaginatively.

When one examines the data from Groups I-IV, one finds that Aggression scores from creative stories correlate  $+ .26$  with Aggression scores from memory stories. Although this  $r$  is in the predicted direction, it is barely significant at the .05 level on a one-tailed test. It is not significantly lower than comparable  $r$ s from Groups I'-IV' and Groups V-VI; and like the other  $r$ s it is obviously far too small to warrant reliable individual diagnosis. For the variable Dependency, the correlation between creative and memory stores for Groups I-IV is insignificantly positive ( $r = + .06$ ).

In short, for both Aggression and Depend-

TABLE 2

CORRELATIONS BETWEEN MOTIVATIONAL SCORES FROM TWO SETS OF TAT STORIES WRITTEN UNDER INSTRUCTIONS TO WRITE CREATIVELY, AND TO DRAW ON MEMORY SOURCES

Group	N	Aggression Dependency	
Memory versus Creative			
I-IV	48	.26*	.06
I'-IV'	36	.37**	-.08
Creative versus Creative			
V-VI	24	.43**	.38*

\*  $p < .05$ ; one-tailed test.

\*\*  $p < .02$ ; one-tailed test.



TABLE 3

INTERNAL CONSISTENCY CORRELATIONS FOR MOTIVATIONAL SCORES WITHIN SETS OF SIX TAT STORIES WRITTEN UNDER INSTRUCTIONS TO WRITE CREATIVELY, AND TO DRAW ON MEMORY SOURCES

Group	N	Aggression		Dependency	
		Creative	Memory	Creative	Memory
I-IV	48	.46***	-.03	.63***	-.01
I'-IV'	36	.38***	.35**	.46***	.26
V-VI	24	-.03 (Session 1)	—	.27 (Session 1)	—
		.11 (Session 2)		.11 (Session 2)	

Note.—Values are Cronbach alphas, corrected by the Spearman-Brown formula.

\*\*  $p < .02$ ; one-tailed test.

\*\*\*  $p < .01$ ; one-tailed test.

ency, the creative-creative correlations are significantly positive, even though too small for individual diagnosis. Intraindividual consistency is also apparent in the creative-memory correlations for the variable Aggression, but for the variable Dependency intraindividual consistency disappears when comparisons are made across the two differing situations: writing from memory sources and writing stories from creative imagination.

How can one interpret the fact that the obtained  $r$ s are low or essentially zero? There are several obvious possible sources of intraindividual inconsistency within the experimental situation. First of all, it may be important that the two scores correlated were based on sets of stories written about a week apart. There are many chances for varying influences to have affected individual subjects during the elapsed time and at either of the two specific occasions when they wrote their stories. Then there is a possibility that a particularly large amount of intraindividual inconsistency is associated with one or the other of the card sets used. Or perhaps the failure to get high correlations between creative and memory stories is due to intraindividual inconsistencies associated with one or the other of the types of instructions used.

Table 3 presents internal consistency coefficients for each of the three combined groups as a function of type of instruction (creative or memory). The values given are Cronbach alphas, computed by the method described in Guilford (1954, p. 385). This type of coefficient, being a function of all

possible split-halves, is independent of the chance goodness of selection of test halves. Between any two data arrays correlated, session number and card set are comparable. Inspection of Table 3 shows that there is no obvious tendency for intraindividual consistency to vary as a function of type of instruction, with the exception that reliability is always zero for scores on memory stories drawn from the pool of stories provided by the experimenter (Groups I-IV).

To examine more minutely the possible role of instructions, a split-half tetrachoric  $r$  was estimated within each session of each of the 10 groups using the method described in Edwards (1954). This yielded 20  $r_t$  values for Aggression, and 20  $r_t$  values for Dependency (data not tabled). Obviously these coefficients are unstable, since each is based on an  $N$  of only 9 or 12 cases. However, the 20 correlations from a given motivational score (e.g., Aggression) can be divided for visual inspection into two arrays, by type of instruction, with card set and session number comparable between the two arrays of  $r_t$  values. When this is done, it appears that a slightly larger number of positive split-half coefficients come from the sets of stories written under creative instructions than from stories written under memory instructions. No significance tests seem to be either warranted or appropriate.

Returning to the possible influence of the card sets, one can divide the 20 split-half  $r_t$  values for a given motivational score into two arrays, associated with Card Sets A and B respectively. Now type of instruction and

session number are comparable between the two arrays of  $r_t$  values. Visual inspection of the two arrays makes it obvious that intraindividual consistency is not a function of card set.

Finally, one can divide the 20 split-half  $r_t$  values for a given motivational score into two arrays, associated with Session 1 and Session 2, respectively. In this division, type of instruction and card set are comparable between the two arrays of  $r_t$  values. Here, too, visual inspection of the two arrays makes it obvious that intraindividual consistency is not a function of the session number at which the stories were written.

#### DISCUSSION

The most striking finding of the study is the low degree of intraindividual consistency found in any of the conditions explored. This finding cannot be laid to unreliable scoring, nor to an arbitrarily unfortunate choice of split halves, in the case of the intrasession correlations.

It appears that drawing memory stories from an artificially prepared pool depresses intraindividual consistency among the memory stories written at a single sitting. However, these coefficients are no lower than some of those based upon sets of creative stories written at a single sitting. When the subjects draw their memory stories from their own reading or other experiences, the degree of intraindividual consistency is not significantly lower than that occurring within most of the sets of creative stories. Thus we cannot reject the clinician's implicit hypothesis that characteristic individual differences influence free choice of memory stories to the same degree as they influence subjects' more genuinely creative stories. The trouble is, of course, that there is so little intraindividual consistency within any of the sets of stories written at a single setting.

Since the intrasession correlations for Aggression are low, and in some cases insignificant, it is surprising to find that all the intersession correlations for this variable reach minimum significance levels. Perhaps the significant intersession  $r_s$  may be due to

greater similarity between card sets than within card sets, but neither our data nor published information permit us to check this idea.

In contrast to the findings on Aggression, the creative-memory correlations for Dependency are insignificant. The unexplainable fact that Aggression and Dependency variables yield dissimilar outcomes implies the need for caution in generalizing from limited research on the TAT. It may turn out that generalizations concerning intraindividual consistency must be quite specific to the scoring variable used.

All in all, the results seem to be a specific case in point for Campbell's (1960) recent statement.

The scientific evidence justifying the introduction of projectives was solely the evidence of the "failure" of the structured approaches and not in the least evidence of the superior validity of the projectives. . . . Now belatedly projective tests are being checked in ways similar to those that invalidated the structured tests, and the evidence for projectives looks even worse (p. 551).

Since clinicians do in fact use the TAT for individual diagnosis, it is not appropriate to rationalize or evade such evidence of weak or nil intraindividual consistency as this study provides. Instead, much more research of the sort presented here should be done, in order to see whether such weak tendencies toward intraindividual consistency are characteristic of various TAT scores for varying card combinations and subjects.

#### REFERENCES

- ARON, BETTY. *A manual for analysis of the Thematic Apperception Test*. Berkeley, Calif.: Willis E. Berg, 1512 Spruce Street, 1949.
- CAMPBELL, D. T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. *Amer. Psychologist*, 1960, 15, 546-553.
- EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- LINDZEY, G. *Projective techniques and cross-cultural research*. New York: Appleton-Century-Crofts, 1961.

(Received August 15, 1962)



## EVIDENCE REGARDING THE EQUIVALENCE OF IPSATIVE AND NORMATIVE PERSONALITY SCALES

ALFRED B. HEILBRUN, JR.

University of Iowa

The equivalence of ipsative and normative personality measures was studied both with regard to interscale correlation and relative validities. Need achievement and nurturance were measured by normative check lists, forced-choice ipsative scales, and ipsative Q sorts, composed of identical items and administered to the same 197 college Ss. The interscale correlations ranged from .43 to .72, the highest correlations being between the normative and Q sort measures. Only the forced-choice ipsative measure of need achievement related significantly to college GPA with ability partialled out, whereas the normative scale of nurturance showed the only reliable relationship with number of charitable activities. The results were interpreted as supporting the use of ipsative measures for normative predictions.

Personality scales fall into one of two broad classes with respect to the nature of the predictions which can logically be generated from them. If the self-ratings or observer-ratings on a given trait are based upon the comparison of that trait with other traits within a single individual, then the scores allow one to predict that this particular individual is more (or less) likely to behave in one fashion than another. The "strengths" of the traits are measured relative only to the strengths of other dimensions of personality within the same individual.

The second class of personality scales includes those which require the rater to specify whether trait-relevant behaviors are or are not characteristic of an individual or, perhaps, how characteristic these may be relative to some reference group. In either case, the scores which a person obtains can be compared across individuals and predictions made regarding the probability that the person will or will not behave in a particular way relative to other people.

Cattell (1944) designated those measures whose scores reflect only the relative strengths of traits within a given individual as *ipsative*, whereas he termed those measures whose scores reflect the strength of traits for a given individual relative to other individuals as *normative*. Cattell felt that scores derived by these two procedures differed in meaning but allowed that "... normative treatment of scores primarily ipsative ... seems legiti-

mate and real" (p. 302). Guilford (1954, p. 528), however, has warned against the use of between-individual differences upon ipsative measures, because there is no single scale for all individuals.

The question of the degree of equivalence between ipsative and normative measures is of practical as well as theoretical significance, since it has become rather common for normative statements to be derived from ipsative measures. The Edwards Personal Preference Schedule (EPPS) (Edwards, 1959) and the various Q sort procedures would be two examples.

As Block (1957) has pointed out, the equivalence of ipsative and normative measures is open to empirical investigation as well as rational argument. He compared observers' ratings of personality obtained from normative rating scales and from ipsative Q sorts and found a remarkably high degree of equivalence (median  $r = .95$  and .79 for the two conditions). Block speculated that the high correspondence between measures was attributable to a normative frame of reference which existed for the expert judges even in an ipsative Q sort procedure. Thus, in the ipsative rating

... the decision between qualitatively different variables as to "salience" is made with reference to a very large body of information and evaluation as to the psychology of human beings (which) ... is based upon some sum or integration of the rater's lifetime experience with and conceptualization of various personalities (pp. 52-53).

Kogan and Fordyce (1962) compared the relationships between three ipsative *Q* sort forms and a normative check list and found the correlations to range between .63 and .70. They concluded that the two types of measures had "practical equivalence."

The present study sought to broaden the evidence regarding the equivalence of ipsative and normative measures in two ways. Two types of ipsative scales were evaluated, whereas Block as well as Kogan and Fordyce investigated only *Q* sorts. Also assessment of equivalence was extended beyond interscale relationships to include the relative validities of ipsative and normative scales.

### METHOD

*Subjects.* The 197 subjects employed in the investigation were volunteers from a large undergraduate class at the University of Iowa. This number included 99 males and 98 females.

*Personality scales.* The two personality traits chosen for study were need achievement and need nurturance. The normative measures of achievement (Ach-N) and nurturance (Nur-N) were selected from the Need Scales (Gough & Heilbrun, in press), which are scored from self-descriptions on the 300-item Adjective Check List (ACL) (Gough & Heilbrun, in press). Evidence has accrued supporting the validity of both scales (Heilbrun, 1958, 1959, 1960, 1962). The correlation between Ach-N and Nur-N is low ( $r = .19$ ), and the item overlap is negligible (3 out of 102 adjectives included in both scales). The ACL is administered to the subject with free-choice, present or absent checking instructions, and the scale *T* scores are based upon the number of adjectives endorsed as self-characteristic which indicate the presence of a given need minus the number of endorsed adjectives which contraindicate this.

Two types of ipsative scales were developed as measures of need achievement and nurturance for the study. One type was a forced-choice measure modeled after the Edwards Personal Preference Schedule to the extent that the subject was presented with items which included two adjectives, each measuring a different need, and was required to select the more characteristic adjective within each pair. The other type was a *Q* sort procedure. The forced-choice measure of achievement (Ach-FC) and nurturance (Nur-FC) utilized the adjectives included in the normative scales described above. Frequency of usage for these adjectives for college students was first estimated in order to control for self-descriptive relevancy as a source of variance in adjective selection. The endorsement base rates for the achievement and nurturance ACL adjectives were estimated from the records of 100

male and 100 female undergraduate subjects not included in the present study. Adjectives from the Ach-N and Nur-N scales were then paired so that each pair included either indicative or contraindicative adjectives, and so that the base rates were as similar for each pair of adjectives as was possible. Different forms of the forced-choice measure were developed for each sex. The matching proved moderately successful in that the average deviation in base rate between paired adjectives was approximately 5% for the male form and 11% for the female form. Both forced-choice measures had 66 adjective pairs, including two pairs keyed to correct for the two overlapping achievement and nurturance adjectives which are scored in opposite directions. The forced-choice measure was administered with instructions for the subject to choose the more characteristic adjective in each pair, omitting none. Scoring proceeded by giving a plus score for each indicative achievement adjective endorsed and a minus credit for each contraindicative achievement adjective endorsed and subtracting the total minus from the total plus. Nur-FC was scored in identical fashion. This procedure requires that each subject will receive the same average score across Ach-FC and Nur-FC which corresponds to the identity of average raw scores on the ipsative EPPS across the 15 variables measured.

The ipsative *Q* sort decks were composed of the 64 Nur-N adjectives and 38 Ach-N adjectives printed on individual cards. The nurturance and achievement decks were administered to the subject separately with instructions to sort them into eight piles as a function of how characteristic the adjective on each was; a rectilinear distribution of cards was required.<sup>1</sup> Scoring for Ach-*Q* and Nur-*Q* was based upon the algebraic summation of plus and minus credits over all cards in a deck. A +7 score was assigned to any card bearing an indicative adjective and placed in the "highly characteristic" category down to a zero if an indicative adjective were put in the highly uncharacteristic pile. Similarly, a -7 was scored if a contraindicative adjective were placed in the highly characteristic pile down to a zero if such an adjective were sorted as "highly uncharacteristic."

It might be noted at this point that since the three types of scales include identical test items and the scale scores were obtained from the same subjects under similar testing conditions, the form of item presentation appears to be the only important scale difference.

<sup>1</sup> A rectilinear distribution of sorts was required because of an earlier finding by Jones (1956) that the average distribution of personality statements for college subjects under free-sort conditions differed significantly from the quasi-normal distribution which is often presumed. Further, inspection of his results suggested that their average much more closely approximated a rectilinear than a normal distribution.



**Criteria.** The criterion for need achievement was level of academic performance at the University of Iowa as measured by cumulative grade-point average with intellectual ability held constant. Ability level was estimated by the composite percentile obtained on the Iowa freshman entrance examination.

The nurturance criterion was one used previously by Heilbrun (1959). The subjects were asked to list by name the "charitable, medical research, rehabilitation, church, or educational activities to which they had contributed time, money, or personal effects within the previous two years." The criterion score was the total number of listed acts (e.g., money contributed to a specified charity) or activities (e.g., "big brother" program at the school for physically handicapped children) which were judged to be both charitable and voluntary by the investigator.

**Procedure.** The subjects were tested in two sessions. In the first session the ACL and the forced-choice measure were administered; the second session included the two Q sorts and the question regarding charitable activities. The order of tests given above was used for all subjects. Median testing time was well below an hour for each session, and the time between sessions was no more than 2 weeks for any subject. In a few cases, the subjects appeared for one testing session and not the other, and the results of their tests appear in only some of the analyses. For this reason and because the entrance examination scores were unavailable for a small number of subjects, the *N*s reported for the various statistical analyses in this study will vary slightly.

## RESULTS

**Interest relationships.** The correlations<sup>2</sup> between the three measures of need achievement and need nurturance are presented in Table 1. All *r*s were positive, of moderate magnitude, and highly significant. The correlations among the achievement scales did not differ reliably, although the correlation between Ach-N and Ach-Q ( $r = .50$ ) was higher than that between the two ipsative scales ( $r = .43$ ). This same finding can be observed for the nurturance measures. The normative scale's *r* with Nur-Q (.72) was higher than was the *r* between ipsative scales (.45), and, in this case, the difference was highly significant as evidenced by a *t* for correlated coefficients (Lindquist, 1940, p. 218) of 5.32 ( $p < .001$ ).

<sup>2</sup> Since there was no basis for expecting the interest relationships or the test-criterion relationships to interact with sex, proportional numbers of male and female subjects were combined for all statistical analyses in the present study.

TABLE 1

CORRELATIONS BETWEEN IPSATIVE AND NORMATIVE SCALES OF NEED ACHIEVEMENT AND NEED NURTURANCE

	Need achievement		Need nurturance	
	Normative	Forced-choice ipsative	Normative	Forced-choice ipsative
Q sort ipsative	.50	.43	.72	.45
Forced-choice ipsative	.43		.50	

Note.—The numbers for the *r*s reported in this table range from 189–193 and include equal proportions of males and females.  $r = .20$  significant at  $p < .01$ .

**Relationships with criteria.** The partial correlations between the various achievement scales and college GPA with ability held constant were as follows: Ach-N,  $r = .07$ ; Ach-FC,  $r = .31$ ; and Ach-Q,  $r = .09$ . The only partial correlation which differed reliably from zero was that between Ach-FC and college performance ( $t = 4.46$ ,  $p < .001$ ). Although none of the partial *r*s differed significantly from each other, the difference between  $r = .31$  for Ach-FC and  $r = .07$  for Ach-N approached significance ( $t$  for correlated coefficients = 1.91,  $p < .06$ ).

The curtailed range of scores on the nurturance criterion (i.e., 82% of the males and 60% of the females reported four or less charitable acts) made correlational analysis inappropriate. Instead, high and low nurturance groups were defined by splitting the male and female criterion score distributions as near as possible to their respective medians and combining the two high and the two low groups. The combined high nurturance group ( $N = 93$ ) included 57 males (charitable acts  $\geq 3$ ) and 36 females (charitable acts  $\geq 5$ ), and the low nurturance group ( $N = 99$ ) was composed of 39 males and 60 females. The mean scores for these two criterion groups for each of the nurturance scales were then compared. These results can be found in Table 2. Only Nur-N related significantly to the criterion.

Since the preceding nurturance analysis did not provide a direct between-scale comparison of validity, one further analysis was conducted. The distribution of scores on

TABLE 2  
COMPARISON OF HIGH AND LOW NURTURANT GROUPS ON IPSATIVE AND  
NORMATIVE NURTURANCE SCALES

Criterion group	Nurturance scale											
	Normative				Forced-choice ipsative				Q sort ipsative			
	N	M	SD	t	N	M	SD	t	N	M	SD	t
High nurturant	91	51.51	10.14	2.29*	93	1.20	12.22	.71	92	98.31	33.00	.50
Low nurturant	98	47.71	12.28		99	2.44	12.58		100	95.87	36.88	

\*  $p < .05$ .

each nurturance scale was divided near the median for each sex and the two high and two low groups were combined. These high-low scale groupings were used in combination with the high-low criterion groupings to determine the proportion of "hits" (e.g., the subject falls in high scale group and high criterion group) and "misses" (e.g., the subject falls in high scale group but low criterion group) afforded by each scale. The three proportions were: Nur-N, 110-79; Nur-FC 98-94; and Nur-Q, 97-95. Comparison of these proportions by chi square tests failed to indicate any significant differences in the proportions, the  $\chi^2$  values ranging from .06 - 1.97.

#### DISCUSSION

The interscale correlations which were obtained in the present study presumably fall below the reliabilities of the individual scales, although only the 10 week test-retest coefficients of the two normative scales were known (Ach-N = .84, Nur-N = .75). Accordingly, it seems safe to assume that each scale will show a higher relationship with itself over time than it will with other concurrent measures of the same variable used in this study. Despite this general caution about scale equivalence, the intertest correlations tend to support the contention that ipsative and normative personality measures were functionally interchangeable in making comparisons between individuals. This is inferred from the fact the correlations between the two ipsative scales were less than those between the ipsative and normative scales, suggesting that the uncommon vari-

ance among scales is primarily attributable to factors other than the ipsative-normative mode of item presentation. This line of evidence must be regarded with some caution, however, because of the nature of construction of the two ipsative measures. The forced-choice measures involved the relative comparison of behaviors representing two needs, whereas the Q sorts involved relative comparisons of behaviors related to a single need. The latter procedure may allow the test-taker to utilize a more normative model in self-evaluation. The correlations reported in Table 1 tend to support this, since the Q sort ipsative versus normative  $r$ s were higher than the forced-choice ipsative versus normative  $r$ s. Thus, it may be important to stipulate the specific nature of the ipsative scale in talking about functional equivalence to normative measures.

The other and, perhaps, more crucial line of evidence which suggests that ipsative measures can be used to make normative statements is that the forced-choice ipsative scale of achievement did in fact correlate significantly with college achievement, ability being held constant. Thus, the self-appraised strength of achievement motivation relative to need nurturance within the individual related to the actual achievement of that individual relative to other individuals. This finding deserves special emphasis since the previously validated normative achievement scale failed to predict academic performance using the same subjects.

The possibility remains that the relative success of Ach-FC was a consequence of the



particular variables which were inadvertently selected for the forced pairing. Achievement in a competitive academic situation may involve an actual element of decision between need achieving and nurturant behavioral alternatives for many students; certainly, attempting to outperform others and extending benefits to others are to some degree incompatible. If so, forcing the subject to choose between achieving and nurturant test alternatives would be consonant with a forced-choice between actual achievement and nurturant social alternatives. The ipsative scale would be matched with an "ipsative" criterion.

The relationships between ipsative and normative ratings reported by Block (1957) (prior to correcting for attenuation) and later by Kogan and Fordyce (1962) were generally higher than the between-scale correlations found in the present study. These lower magnitudes are consistent with Block's hypothesis that the high correspondence between ipsative and normative ratings in his study was due to a broader frame of reference used by the judges when traits were considered relatively within the individual. Since it can be presumed that the trained psychologists employed by Block share a greater "body of information . . . as to the psychology of human beings" and, thus, a more applicable frame of reference than college adolescents, lower coefficients would be expected in the current study. In any case, the present results, taken in conjunction with earlier findings by Block and by Kogan and Fordyce, provide a substantial case for the equivalence of ipsative

and normative personality measures for making normative predictions.

## REFERENCES

- BLOCK, J. A. A comparison between ipsative and normative ratings of personality. *J. abnorm. soc. Psychol.*, 1957, 54, 50-54.
- CATTELL, R. B. Psychological measurement: Ipsative, normative, and interactive. *Psychol. Rev.*, 1944, 51, 292-303.
- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1959.
- GOUGH, H. G., & HEILBRUN, A. B. *Joint manual for the Adjective Check List and the Need Scales for the ACL*. Palo Alto, Calif.: Consulting Psychologists Press, in press.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- HEILBRUN, A. B. Relationships between the Adjective Check List, Personal Preference Schedule and desirability factors under varying defensiveness conditions. *J. clin. Psychol.*, 1958, 14, 283-287.
- HEILBRUN, A. B. Validation of a need scaling technique for the Adjective Check List. *J. consult. Psychol.*, 1959, 23, 347-351.
- HEILBRUN, A. B. Personality differences between adjusted and maladjusted college students. *J. appl. Psychol.*, 1960, 44, 341-346.
- HEILBRUN, A. B. Prediction of first year college drop-out, using ACL Need Scales. *J. counsel. Psychol.*, 1962, 9, 58-63.
- JONES, A. Distributions of traits in current Q sort methodology. *J. abnorm. soc. Psychol.*, 1956, 53, 90-95.
- KOGAN, W. S., & FORDYCE, W. E. The control of social desirability: A comparison of three different Q sorts and a check list, all composed of the same items. *J. consult. Psychol.*, 1962, 26, 26-30.
- LINDQUIST, E. F. *Statistical analysis in educational research*. Cambridge, Mass.: Riverside Press, 1940.

(Received August 20, 1962)

## HYPNOTIC SUSCEPTIBILITY AND MMPI PROFILES

ROBERT E. SCHULMAN<sup>1</sup> AND PERRY LONDON

*University of Illinois*

4 groups totaling 87 undergraduate females of relatively different hypnotic susceptibility as determined by Form A of the Stanford Scale of Hypnotic Susceptibility completed the questionnaire form of the MMPI. The most susceptible group scored significantly lower than the others on the Pd scale ( $p < .01$ ), but no other significant differences were found between groups. Mean scores of all susceptibility groups fell within the normal range, suggesting that hypnotic susceptibility is not grossly related to psychopathology.

The recent revival of interest in formal research on hypnosis and hypnotic susceptibility has been accompanied by renewed interest in the personality characteristics and adjustive states of the hypnotically susceptible individual. The notion was popular among nineteenth century psychiatrists, such as Janet, that susceptibility to hypnosis was a trait found primarily in hysterical neurotics, and to this day psychoanalytic theorists have tended to describe the ability to enter hypnotic states in terms of temporarily regressive (Schilder, 1956), if not permanently pathological behavior (Fenichel, 1945).

Theorizing in this area was almost entirely unhindered by the results of empirical research until Wells' (1931) early work which attempted to relate hypnotic susceptibility to various personality traits. His results however were generally contradictory and unreliable. White (1937) used the Thematic Apperception Test to demonstrate that susceptibility to hypnosis can be predicted on the basis of an individual's social attitudes and attitudes towards the phenomenon itself, a finding later confirmed by Rosenzweig and Sarason (1942).

The Rorschach has also been used to study susceptibility (Brenman & Reichard, 1943; Sarbin & Madon, 1942; Steisel, 1952), but results of these studies have been conflicting and ambiguous. It is a fair conclusion that projective tests have by and large been unsuccessful in identifying clinical and personality correlates of hypnosis.

A number of recent studies have found low

but significant correlations between some scales of inventory tests such as the Minnesota Multiphasic Personality Inventory, Cattell 16 PF, and California Personality Inventory (CPI) with volunteering as subjects for hypnosis and similar behaviors (London, 1961; London, Cooper, & Johnson, 1962), but these studies did not specifically examine differences between hypnotically susceptible and unsusceptible subjects. Hilgard, Weitzenhoffer, Landes, and Moore (1961) have used the CPI for the latter purpose, but generally with inconclusive results.

With the exception of the work of Weitzenhoffer (1958), Hilgard et al. (1961), and Hilgard and Lauer (1962), most of the studies which used projective tests and factor analyzed personality inventories to study susceptibility were methodologically limited by such details as small sample size, etc. It is worth considering, however, that their results would in any case have to be limited to test-related statements about personality characteristics whose clinical implications were at best indirect.

Among existing tests, only the MMPI is empirically and directly derived from norms obtained on relevant pathological populations. Three studies related to susceptibility have reported using the MMPI (Faw & Wilcox, 1958; Sarbin, 1950; Sector, 1961). Their results appear generally unimpressive, but may be obscured by ambiguities in reporting criteria and results as well as by the use of small and biased samples.

The present research was concerned with examining MMPI scale and profile differences for four groups of females of different relative hypnotic susceptibility.

<sup>1</sup> Presently a Postdoctoral Fellow in Adult Clinical Psychology at the Menninger Foundation.



## METHOD

The *Stanford Hypnotic Susceptibility Scale*, Form A (Weitzenhoffer & Hilgard, 1959), was individually administered to 87 female undergraduates who had volunteered to participate in a hypnotic experiment. The questionnaire form of the MMPI was subsequently completed by each subject as follows:

Subjects were classified into four groups of relative hypnotic susceptibility on the basis of their *Stanford Scale* scores. There were 16 subjects in the High (10-12) scoring group (*H*), 19 in the Medium (8-9) group (*M*), 33 subjects obtained scores placing them in the Low (5-7) group (*L*), and 19 fell in the relatively Unsusceptible (0-4) group (*U*).

None of the subjects were informed of their precise susceptibility ratings, but members of the two relatively high susceptibility groups, *H* and *M*, as well as of the relatively unsusceptible group, *U*, were immediately assigned to another experiment. This experiment was finished within 3 weeks after the *Stanford Scale* administration, and the subjects then completed the MMPI. Members of the *L* group were not assigned to another experiment, however, and completed the MMPI within a few days of the *Stanford Scale*.

## RESULTS

Inspection of the average profiles and means (Table 1) of the four groups plainly suggests that they are all more alike than different.

A simple four-way analysis of variance (Walker & Lev, 1953) was performed separately on each MMPI scale. The *Pd* scale showed significant mean differences between groups ( $p < .01$ ), as did the *Pt* scale ( $p < .05$ ) (Table 2). In both instances the

TABLE 1

MMPI MEAN *T* SCORES

Scale	Susceptibility group			
	Unsusceptible (0-4)	Medium (8-9)	High (10-12)	Low (5-7)
<i>L</i>	46.53	43.63	47.00	46.00
<i>F</i>	50.84	53.68	53.69	53.82
<i>K</i>	56.58	57.05	56.19	54.33
<i>Hs</i>	49.63	49.74	52.12	48.91
<i>D</i>	48.47	49.74	47.12	51.48
<i>Hy</i>	56.11	53.32	54.69	56.45
<i>Pd</i>	57.16	59.16	53.94	59.67
<i>Mf</i>	46.42	46.11	49.44	44.12
<i>Pa</i>	55.74	53.42	54.75	57.27
<i>Pt</i>	54.42	54.95	51.12	58.58
<i>Sc</i>	55.79	57.42	55.12	60.03
<i>Ma</i>	55.79	58.58	58.56	60.82
<i>Si</i>	47.32	50.95	46.31	50.21

*H* group was significantly lower than the other three groups.

Since the sequence of conditions for the *L* group differed from that of the three groups who had participated in a separate experiment, a simple three-way analysis of variance was completed for the three groups who had taken the MMPI 3 weeks after having been administered the *Stanford Scale*. The results were generally the same as those for the four-way analysis except that the *Pt* scale difference did not occur; *Pd* scale differences, however, were still significant (Table 3).

Each profile was then scored for high and

TABLE 2  
ANALYSIS OF VARIANCE OF FOUR SUSCEPTIBILITY GROUPS

Variable	Total $\frac{df}{87}$	Between $\frac{df}{3}$	Within $\frac{df}{83}$	<i>F</i>
<i>L</i>	2,777	125		
<i>F</i>	3,519	124	2,652	1.30
<i>K</i>	5,208	113	3,395	1.00
<i>Hs</i> + .5 <i>K</i>	4,162	112	5,095	—
<i>D</i>	7,332	238	4,050	—
<i>Hy</i>	6,197	135	7,094	—
<i>Pd</i> + .4 <i>K</i>	7,384	965	6,062	—
<i>Mf</i>	7,137	292	6,419	4.16*
<i>Pa</i>	7,391	194	6,845	1.18
<i>Pt</i> + 1 <i>K</i>	7,100	645	7,197	—
<i>Sc</i> + 1 <i>K</i>	6,529	355	6,455	2.76**
<i>Ma</i> + .2 <i>K</i>	11,740	308	6,174	1.59
<i>Si</i>	8,348	288	11,432	—
			8,060	—

\*  $p < .01$ .\*\*  $p < .05$ .

TABLE 3  
ANALYSIS OF VARIANCE OF THREE SUSCEPTIBILITY GROUPS

Variable	Total $\frac{df}{54}$	Between $\frac{df}{2}$	Within $\frac{df}{51}$	F
<i>L</i>	1,682	122	1,560	1.99
<i>F</i>	2,228	98	2,130	1.17
<i>K</i>	3,061	6	3,055	—
<i>Hs</i> + .5 <i>K</i>	2,798	67	2,731	—
<i>D</i>	4,348	59	4,289	—
<i>Hy</i>	2,902	73	2,829	—
<i>Pd</i> + .4 <i>K</i>	4,067	810	3,257	6.34*
<i>Mf</i>	4,361	107	4,254	—
<i>Pa</i>	3,663	51	3,612	—
<i>Pt</i> + 1 <i>K</i>	3,496	145	3,351	1.10
<i>Sc</i> + 1 <i>K</i>	3,170	49	3,121	—
<i>Ma</i> + 2 <i>K</i>	8,520	95	8,425	—
<i>Si</i>	4,612	214	4,398	1.24

\*  $p < .01$ .

low point codes, a procedure which could reveal patterns of differences between groups regardless of the similarities of scale scores. Examination of these codes did not reveal any meaningful groupings relative to the four levels of susceptibility.

### DISCUSSION

The analysis of the data indicates that only one of the MMPI's clinical scales seems to distinguish consistently between groups which differ in hypnotic susceptibility. The *Pd* scores of the most susceptible group (*H*) were significantly lower than those of any other group in both analyses. The absolute differences between them even for this scale, however, were small ones. Taking these differences at face value, one might allege post hoc that very susceptible people are less aggressive, more compliant, and less negative about relations with authority figures than are others. But their very willingness to respond to the hypnotic situation communicates this idea more effectively than the MMPI, and the size of the difference on *Pd* is too small to argue for any great difference in either clinical status or behavior.

The scale performances of members of all groups are essentially normal ones, with virtually all means falling within a single standard deviation of the mean of the Normal standardization group. Since the MMPI is in general not very adept at identifying subtle

personality differences within the normal range, it is hardly surprising that it is insensitive to possible differences between groups of this sample. It is of interest also that our results offer no support whatsoever for the connections between *Hy* and susceptibility previously obtained with smaller samples. All told, the conclusion seems in order that susceptibility to hypnosis can best be viewed as a trait falling within the normal range of behavior rather than representing hysteria or any other gross psychopathology.

A more general conclusion concerning both the present study and the host of work preceding it may be worth noting as well: There may indeed be personality traits which distinguish persons of relatively different degrees of hypnotic susceptibility, and these traits may be well worth discovering; but it seems quite clear that they are not going to be discovered by any of our existing gross personality inventories, whether the factor analyzed ones used by previous workers or the empirically derived one used here. Neither our old Kraepelinian nor current construct categories seem very relevant to this trait. It is time to stop doing studies like this one and to seek a fresh approach.

### REFERENCES

- BRENMAN, M., & REICHARD, S. Use of the Rorschach test in the prediction of hypnotizability. *Bull. Menninger Clin.*, 1943, 11, 60-66.



- FAW, V., & WILCOX, W. W. Personality characteristics of susceptible hypnotic subjects. *J. clin. exp. Hypn.*, 1958, 6, 83-94.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- HILGARD, E. R., & LAUER, LILLIAN W. Lack of correlation between the California Psychological Inventory and hypnotic susceptibility. *J. consult. Psychol.*, 1962, 26, 331-335.
- HILGARD, E. R., WEITZENHOFFER, A. M., LANDES, J., & MOORE, ROSEMARIE K. The distribution of susceptibility to hypnosis in a student population: A study using the Stanford Hypnotic Susceptibility Scale. *Psychol. Monogr.*, 1961, 75(8, Whole No. 512).
- LONDON, P. Subject characteristics in hypnosis research: I. A survey of experience, interest, and opinion. *Int. J. clin. exp. Hypn.*, 1961, 9, 151-161.
- LONDON, P., COOPER, L. M., & JOHNSON, H. J. Subject characteristics in hypnosis research: II. Attitudes toward hypnosis, volunteer status, and personality measures. III. Some correlates of Hypnotic Susceptibility. *Int. J. clin. exp. Hypn.*, 1962, 10, 13-21.
- ROSENZWEIG, S., & SARASON, S. An experimental study of the triadic hypothesis in relation to frustration, ego defense and hypnotic ability. *Charac. Pers.*, 1942, 11, 150-165.
- SARBIN, T. R. Contributions to role-taking theory: I. Hypnotic behavior. *Psychol. Rev.*, 1950, 57, 255-270.
- SARBIN, T. R., & MADON, L. W. Predicting the depth of hypnosis by means of the Rorschach test. *Amer. J. Orthopsychiat.*, 1942, 12, 268-271.
- SCHILDER, P. *The nature of hypnosis*. New York: International Univer. Press, 1956.
- SECTER, I. I. Personality factors of the MMPI and hypnotizability. *Amer. J. clin. Hypn.*, 1961, 3, 185-188.
- STEISEL, I. M. The Rorschach test and suggestibility. *J. abnorm. soc. Psychol.*, 1952, 47, 608-614.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
- WEITZENHOFFER, A. M. Personality and hypnotic susceptibility. *Amer. J. clin. Hypn.*, 1958, 1, 79-82.
- WEITZENHOFFER, A. M., & HILGARD, E. R. Stanford Hypnotic Susceptibility Scale. Palo Alto, Calif.: Consulting Psychologists Press, 1959.
- WELLS, W. R. Hypnotizability vs. suggestibility. *J. abnorm. soc. Psychol.*, 1931, 25, 436-449.
- WHITE, R. W. Prediction of hypnotic susceptibility from a knowledge of subjects' attitudes. *J. Psychol.*, 1937, 3, 265-277.

(Received August 22, 1962)

## REVISION OF THE MMPI K CORRECTION PROCEDURE FOR IMPROVED DETECTION OF MALADJUSTMENT IN A NORMAL COLLEGE POPULATION<sup>1</sup>

ALFRED B. HEILBRUN, JR.<sup>2</sup>

University of Iowa

The problem investigated was whether a better system of K weighting could be developed for the 10 MMPI clinical scales to enhance their usefulness as measures of adjustment level within a grossly normal college population. 2-group discriminant analysis was used to determine the K value for each scale to maximize discrimination between maladjusted and adjusted college Ss, the sexes being considered separately. The revised weighting system differed greatly from the standard system; major differences were in negative weighting of Hy and deletion of weights from Hs, Pd, and Ma. Lesser changes in Pt and Sc weights were obtained, and D, Mf, Pa, and Si continued to be unweighted. Cross-validation of revised weights was shown for maladjusted college Ss and seriously maladjusted psychopathic hospital cases.

The Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1951) is probably the most extensively used and validated psychometric device available for diagnosing psychopathology, yet for several years evidence has accrued to suggest that the MMPI has been inefficiently applied to the task of detecting less severe maladjustment within a grossly normal population. In the early stages of development of the test, it was found that certain of the clinical scales showed an improvement in discrimination between hospitalized patients and an unselected normal group when a scale measuring test-taking attitudes was incorporated as a suppressor variable (McKinley, Hathaway, & Meehl, 1956). The K scale was developed as a measure of test-taking defensiveness largely by finding MMPI items which distinguished between normals and hospitalized psychiatric patients who showed normal test profiles and elevated scores on the Lie scale, the assumption being that the latter group were portraying themselves in an unduly favorable light. Empirically determined K weights were then determined to discriminate optimally between new

samples of diagnosed patients and normals. Five clinical scales were corrected by adding from .2 up to 1.0 of the raw K scale score, consistent with the logic that higher K scores mean greater guardedness which, in turn, is associated with unduly depressed (and healthier) profiles on the MMPI. McKinley et al. (1956) warned, however, that

... these weights are optimal, within our sample, for the differentiation of largely inpatient psychiatric cases of full-blown psychoneurosis and psychosis from a general Minnesota "normal" group. For other clinical purposes it is possible that other  $\lambda$  values would be more appropriate. Thus, it seems likely that for the best separation of "maladjusted normals," such as those which abound in a college counseling bureau and would be formally diagnosed in a psychiatric clinic as *simple adult maladjustment*, other weights might be better (p. 115).

Since the person who obtains a high score on K does so by denying personal inadequacies, the significance of a high score obtained from a hospitalized psychiatric patient and from a normal person making an adjustment in the community might be expected to differ. In the former case, *prima facie* evidence suggests a defensive distortion of fact, whereas in the latter case there is a considerably higher probability that the person is giving a factual account of his psychological adequacies. Research evidence has been presented to substantiate the hypothesis that the K factor is positively related to psychological strengths in normal populations.

<sup>1</sup> This investigation was supported by a research grant from the Graduate College, State University of Iowa.

<sup>2</sup> The author wishes to thank Dee Norton for his valuable assistance in the planning and conduct of the statistical analyses of the data.



Block and Thomas (1955) and Berger (1955) found a positive relationship between  $K$  scores and degree of self-acceptance. A picture of the high  $K$  normal as poised and comfortable in social situations emerged from studies at the Institute of Personality Assessment and Research (Block & Bailey, 1955; Gough, McKee, & Yandell, 1955).

Smith (1959) reported a significant *negative* correlation of .39 between  $K$  scores and judgments of defensiveness for his normal subjects; further, higher  $K$  scorers were rated as more insightful. Heilbrun (1961) tested the hypothesis that  $K$  is a measure of psychological health in a grossly normal population and found some support in the case of college females where an adjusted group scored higher on the  $K$  scale than a counseling service maladjusted group. He further demonstrated that the  $K$  scale is more highly correlated with test-taking defensiveness for both male and female maladjusted subjects within a normal college population than for their adjusted counterparts. Nakamura (1960) found that presumably more maladjusted disciplinary cases scored significantly higher on  $K$  when retested as part of an evaluation which might culminate in probation or suspension from school relative to test-retest comparisons for college non-disciplinary controls.

The present study sought to determine empirically the  $K$  corrections for the 10 clinical MMPI scales which would maximize the discrimination between adjusted and maladjusted persons in a college population. The importance of improving the MMPI's ability to distinguish maladjusted persons within a college population goes beyond the clinical benefits to be obtained. Considerable research into the nature of behavioral maladjustment is conducted in college settings, much of it employing the MMPI as the measure of maladjustment. Obviously, the more precisely the psychometric instrument can measure degree and kind of maladjustment within this type of population, the more likely the research will provide lawful relationships.

#### METHOD

*Subjects and procedure.* The study was separated into two phases. The purpose of the initial phase was to determine the optimal amount of the raw  $K$  score to be added to or subtracted from the

raw scores of the 10 MMPI clinical scales to maximize the discrimination on each scale between maladjusted and adjusted subjects in a college population. Maladjustment was defined as having sought help at the University Counseling Service, University of Iowa, for personal adjustment problems. Since the contribution of  $K$  was to be investigated separately by sex, groups of 50 undergraduate male and 50 undergraduate female clients were included in the initial samples. Adjusted, in turn, was defined as not having sought help for any type of problem at the Counseling Service up until the time of data collection for the study. In order to specify more clearly the type of college population to which the results of the study could be generalized, a base rate for occurrence of maladjustment was estimated and the appropriate ratio of adjusted to maladjusted subjects reflected in the size of the respective groups. It was estimated that 10% of college undergraduates experience psychological difficulties of sufficient intensity to warrant personal adjustment counseling. Accordingly, 450 male and 450 female undergraduates were included in the adjusted groups, maintaining a nine-to-one ratio of adjusted to maladjusted subjects.

The second phase of the study included two independent cross-validations of the optimal  $K$  weights obtained in the initial phase. The first cross-validation was a direct replication of the initial phase conditions. MMPI records of 30 male and 30 female undergraduates, seen at the Counseling Service because of personal adjustment difficulties, were compared with those obtained from 270 male and 270 female adjusted undergraduates. The hypothesis tested was whether the 10 clinical scales, using the optimal  $K$  weights obtained in the initial phase, differentiated these two groups more effectively than the clinical scales as standardly weighted.

In addition, the generality of the initial findings was tested by comparing the newly developed  $K$  weighting system to the standard system in the discrimination between more seriously maladjusted college groups and adjusted college students. The maladjusted subjects in this cross-validation were patients at the Psychopathic Hospital, University of Iowa, which specialized in short-term treatment of acutely disturbed psychiatric cases. Fifty male and 50 female patients (about 85% of whom were maladjusted) constituted the more seriously maladjusted groups, all of these being unselected as to psychiatric diagnosis. The adjusted subjects in this comparison were 450 male and 450 female undergraduates. These adjusted samples included the 270 males and 270 females in the other cross-validating comparison but did not overlap with the 900 adjusted subjects employed in the initial phase of the study.

*Test administration.* The MMPI records for 1960 of the 2,060 subjects included in the study were obtained under group testing conditions just prior to the beginning of their freshman year at the university. The exceptions were the 100 records obtained from the Psychopathic Hospital where MMPIs were individually administered as part of a diagnostic battery. Adjusted and maladjusted

groups in the initial and first cross-validation phase are closely matched for age (about 18 years) and education (12 years). No attempt was made to match the Psychopathic Hospital samples for mean age and education with the adjusted comparison groups. The only restrictions imposed for the seriously maladjusted subjects were that they be relatively young (<39 years of age) and that they had at any time attended a college.

The MMPI records for the 1,800 adjusted subjects were drawn from the 1959 freshman class. The MMPIs for the 160 Counseling Service maladjusted subjects were obtained as part of the freshman testing program in all years from 1958-1961, inclusive.

## RESULTS

### Initial Phase

Two-group discriminant analysis was performed to determine the multiple regression formula which would best predict the dichotomous criterion of adjustment versus maladjustment for each raw score combination of  $K$  with each of the 10 MMPI clinical scales. The multiple biserial correlations between the optimally combined raw scores and adjustment level were compared with the biserial correlations between the same scales as standardly scored in  $T$  units, with the traditional  $K$  correlations included on five scales. Two rules were followed in deciding whether a revision in  $K$  weights was warranted.

(a) If the  $K$  scale was assigned a statistically significant beta weight in the multiple

regression formula (i.e., made a contribution to prediction beyond that contributed by the MMPI clinical scale) and the multiple  $R$  was significantly greater than the  $r$  between the standardly scored scale and the adjustment criterion, the new optimal  $K$  weight for that scale was adopted.

(b) If the  $K$  scale was not assigned a statistically significant beta weight in the multiple regression formula and the standardly scored scale did not have a significantly higher  $r$  with the criterion than did the  $K$  uncorrected raw scale score, no  $K$  correction was adopted for that scale.

The results of the analyses of the initial phase data are given in Table 1. The only scales for which  $K$  corrections significantly improved the discrimination of adjusted and maladjusted college subjects relative to raw scale scores alone were  $Hy$ ,  $Pt$ , and  $Sc$ , these results holding true for both males and females.

The differences in correlations with the adjustment criterion between the standardly scored clinical scales and either the new  $K$ -corrected  $Hy$ ,  $Pt$ , and  $Sc$  scales or the raw scores on the seven remaining scales were then tested, and the  $F$  values are reported in the final two columns of Table 1. It can be seen that the three scales with revised  $K$  weights correlate significantly higher with the adjustment criterion than do these same scales as

TABLE 1

COMPARISON OF OPTIMAL AND STANDARD  $K$  SCALE WEIGHTS FOR THE 10 MMPI CLINICAL SCALES IN DIFFERENTIATING LEVELS OF ADJUSTMENT (CRITERION) WITHIN MALE (M) AND FEMALE (F) COLLEGE POPULATIONS

MMPI scales	$r$ $K$ versus scales		$r$ $T$ scores versus criterion (1)		$r$ Raw scores versus criterion (2)		$R$ Raw scores + $K$ versus criterion (3)		Beta for raw scores		Beta for $K$		F Values $r$ (1) versus $r$ (2) or $R$ (3)	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
$Hs$	-.43	-.41	.09	.16	.17	.22	.17	.22	.17**	.23**	.00	.02	12.67**	11.29**
$D$	-.22	-.23	.30	.30	.30	.29	.30	.29	.29**	.29**	-.01	-.01	2.95	2.26
$Hy$	-.28	-.25	.12	.22	.12	.22	.17	.25	.16**	.25**	-.12*	-.14**	6.65*	8.71**
$Pd$	-.29	-.27	.18	.27	.17	.31	.17	.31	.16**	.31**	-.03	.01	.89	10.77**
$Mf$	-.20	-.05	.23	.05	.23	.04	.23	.08	.22**	.04	-.03	.07	.24	.26
$Pa$	-.13	-.04	.11	.18	.11	.18	.12	.19	.10	.18**	-.06	-.06	.67	.00
$Pt$	-.70	-.69	.25	.24	.28	.24	.33	.27	.45**	.35**	.24**	.17**	25.88**	8.34**
$Sc$	-.59	-.60	.29	.26	.29	.27	.32	.30	.38**	.36**	.15**	.15**	7.38**	11.57**
$Ma$	-.45	-.44	.06	.11	.08	.12	.09	.12	.06	.10	-.04	-.03	1.93	1.00
$Si$	-.49	-.49	.16	.06	.16	.11	.16	.11	.16**	.10	.01	-.02	.33	4.39*

$Hs$ =hypochondriasis,  $D$ =depression,  $Hy$ =hysteria,  $Pd$ =psychopathic deviate,  $Mf$ =masculinity-femininity,  $Pa$ =paranoia,  $Pt$ =psychasthenia,  $Sc$ =schizophrenia,  $Ma$ =hypomania,  $Si$ =social introversion.  
 \* Standardly employed scales scored in  $T$  score units.  
 \*  $p < .05$ .  
 \*\*  $p < .01$ .



TABLE 2—RELATIVE PREDICTION OF ADJUSTMENT STATUS IN TWO COLLEGE SAMPLES BY MEANS OF THE MMPI CLINICAL SCALES WITH REVISED AND STANDARD *K* CORRECTION PROCEDURES

MMPI scale	Males				Females			
	Counseling service versus normals (N = 300)		Psychopathic hospital versus normals (N = 500)		Counseling service versus normals (N = 300)		Psychopathic hospital versus normals (N = 500)	
	Revised scale <i>r</i>	Stand-ard scale <i>r</i>	F for differ-ence	F for differ-ence	Revised scale <i>r</i>	Stand-ard scale <i>r</i>	F for differ-ence	F for differ-ence
<i>Hs</i>	.38	.28	22.40**	7.03**	.01	.04	.36	16.17**
<i>D</i>	.31	.17	21.64**	4.12*	.11	.13	1.00	.79
<i>Hy</i>	.29	.26	4.76*	4.29*	.01	.04	.57	50.87**
<i>Pd</i>	.20	.15	5.15*	.33	.09	.05	1.71	30.69**
<i>Mf</i>	.17	.17	.12	1.57	.04	.10	2.76	.10
<i>Pa</i>	.12	.14	1.52	6.71**	.06	.14	4.67*	33.76**
<i>Pt</i>	.25	.20	6.93**	9.67**	.10	.07	1.47	21.56**
<i>Sc</i>	.30	.24	9.51**	8.19**	.16	.13	2.10	21.84**
<i>Ma</i>	.04	.02	.36	.20	.07	.05	.06	6.19*
<i>Si</i>	.16	.16	.09	1.86	.17	.18	.31	.39

Note.—With  $N = 300$ ,  $r = .12$ , and .15 significant at  $p < .05$  and .01, respectively; with  $N = 500$ ,  $r = .09$ , and .12 significant at  $p < .05$  and .01, respectively.  
 \*\*  $p < .01$ .

currently scored. In no case did the standardly scored scale correlate reliably higher than did the simple raw score on the remaining seven scales, and in some instances raw scores related better to the criterion than did the scales with traditional *K* weights.

The final stage of the data analysis in the initial phase was to apply the two rules for revision of *K* weights so that optimally corrected clinical scales for these samples of college subjects could be tested by cross-validation. Three scales met the requirements for Rule 1 for both males and females. The optimal weights for these scales determined from the regression formula and rounded to the nearest tenth were as follows:

	Males	Females
<i>Hy</i>	-.7 <i>K</i>	-.5 <i>K</i>
<i>Pt</i>	1.0 <i>K</i>	.8 <i>K</i>
<i>Sc</i>	.7 <i>K</i>	.7 <i>K</i>

Application of Rule 2 negated the use of any *K* correction for *Hs*, *D*, *Pd*, *Mf*, *Pa*, *Ma*, and *Si* for either male or female samples.

#### Cross-Validation Phase

The results of cross-validation of the revised *K* weighting system in both counseling service and psychiatric hospital discrimination situation are given in Table 2. If revision were effective it would be expected that for those scales where *K* weights were newly constituted (*Hy*) or altered to some value other than zero (*Pt*, *Sc*), prediction of level of adjustment should be significantly better than for the same scales as standardly weighted. Further, for those scales where standard *K* weights were dropped in the revised system (*Hs*, *Pd*, *Ma*) prediction from raw scores should be as good as from standardly weighted scores. Inspection of Table 2 shows that 20 of the 24 comparisons<sup>3</sup> involving the six revised scales, these expectations were fulfilled.

For the four scales which received *K* correction in neither the standard nor the revised system (*D*, *Mf*, *Pa*, *Si*), 13 of the 16 comparisons showed no difference between raw

<sup>3</sup>  $F = \frac{r_1 - r_2/p - 1}{(1 - r_1^2)/N - p - 1}$  where  $p$  (number of variables) = 2,  $N$  (total number of subjects) = 500, and  $df = 1,497$ .

and  $T$  score prediction or that the scale raw scores were more highly correlated. The  $Pa$  scale tended to perform more effectively using the standard  $T$  scores, whereas  $D$  was a more effective predictor using raw scores.

### DISCUSSION

Since the revised system for applying  $K$  weights to the MMPI clinical scales held up rather well upon cross-validation, its use when the MMPI is employed in a college population appears warranted. Needless to say, before any established psychometric technique is replaced by another, more evidence than that offered by a single study is necessary. Several questions still remain unanswered. For one, will the optimal weights derived from students within a single university prove useful in other colleges? What relevant evidence there is (Black, 1956; Goodstein, 1956) suggests that when the modal MMPI profiles for samples of male and female college students from diverse schools are compared, no significant regional differences are found. This might indicate that over large groups optimal  $K$  weights would also be homogeneous from school to school. In any case, if optimal weighting should turn out to be specific to discrete school populations, this would be equally problematic for any universal system including the traditional one. Revised norms for the 10 MMPI clinical scales incorporating the new  $K$  weights are available to those who wish to use and further evaluate the revised correction system.<sup>4</sup>

Another question which remains is whether improved discrimination between maladjusted and adjusted students upon a given scale has any bearing upon the usefulness of that scale as a measure of some specific psychiatric disorder. For example, does the change from .4  $K$  to zero weighting on the  $Hs$  scale for a college population mean that we can make more refined inferences regarding bodily over-concern from scores on that scale or does the changed weight allow only better prediction regarding "degree of adjustment"? Since there is only a limited degree of diagnostic relevance to such terms as "hypochondriacal," "paranoid," "schizophrenic," etc., when the behavioral problems of grossly normal ado-

lescents are considered, cross-validation of the revised scale weights within counseling service samples is not good evidence that the psychiatric meanings of the individual scales have been more clearly specified. However, the revised weights were also cross-validated within a sample of psychiatrically disturbed, college educated subjects for whom such diagnostic classes have more relevance. Accordingly, somewhat more confidence seems warranted in the assumption that the optimal weights not only permit greater precision of assignment to broad adjustment classes but also a more accurate estimate of the specific nature of the psychopathology as given by scores on the individual scales.

### REFERENCES

- BERGER, E. M. Relationships among acceptance of self, acceptance of others, and MMPI scores. *J. counsel. Psychol.*, 1955, 2, 279-283.
- BLACK, J. D. MMPI results for fifteen groups of female college students. In G. S. Welsh & W. G. Dahlstrom, *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- BLOCK, J., & BAILEY, D. Q-sort item analyses of a number of MMPI scales. Technical Memorandum, May 1955, Officer Education Research Laboratory, OERL-TM-55-7.
- BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, 51, 254-259.
- GOODSTEIN, L. D. Regional differences in MMPI responses among male college students. In G. S. Welsh & W. G. Dahlstrom, *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- GOUGH, H. G., MCKEE, M. G., & YANDELL, R. J. Adjective check list analyses of a number of selected psychometric and assessment variables. Technical Memorandum, May 1955, Officer Education Research Laboratory, OERL-TM-55-10.
- HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) New York: Psychological Corporation, 1951.
- HEILBRUN, A. B. The psychological significance of the MMPI  $K$  scale in a normal population. *J. consult. Psychol.*, 1961, 25, 486-491.
- MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The  $K$  scale. In G. S. Welsh & W. G. Dahlstrom, *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- NAKAMURA, C. Y. Validity of the  $K$  scale in college counseling. *J. counsel. Psychol.*, 1960, 7, 108-115.
- SMITH, E. E. Defensiveness, insight and the  $K$  scale. *J. consult. Psychol.*, 1959, 23, 275-277.

(Received August 29, 1962)

<sup>4</sup> Copies of the college norms and the conversion table from raw  $K$  scores to specific weights are available without charge from the author.



## RELATIVE IMPORTANCE OF TEST ITEM CONTENT<sup>1</sup>

WARREN T. NORMAN

University of Michigan

The "Deviation Hypothesis" and some related notions presented by Berg including the contention that test item content is unimportant in personality measurement are critically examined. Some previous findings cited in support of these formulations are briefly reviewed. 2 additional sets of data comparing the effectiveness of test stimuli with different degrees of content relevance are presented. The results clearly indicate that for the criteria and subject populations considered, there exist marked differences in the validities obtainable from different classes of test stimuli—those with the highest degree of judged content relevance producing the most satisfactory results. The theoretical or explanatory value ascribed to the Deviation Hypothesis is called into question.

Irwin Berg (1959, 1961) has argued for the unimportance of test item content in the development of measures to identify what he calls "deviant subgroups" in the population. This proposition, "... that particular content of objective personality and interest tests is unimportant . . . [that] virtually any content of any sense modality should be suitable . . ." (Berg, 1959, p. 86) is actually presented as a corollary to a more general principle which Berg (1955, 1957) has dubbed the *Deviation Hypothesis*.

In the sequel, I propose first to analyze critically Berg's formulation of this principle and its derivatives. I will then cite (but only briefly) some of the data he offers in support of his position. I will next present some new data which I believe bear on Berg's thesis and will argue that, at the very least, the Deviation Hypothesis and its attendant notions have been overstated. Finally, I will contend that the *theoretical* or *explanatory value* which Berg has ascribed to the Deviation Hypothesis is at best trivial, and that at worst, this position leads to an orientation toward personality assessment of a most militantly atheoretical variety.

<sup>1</sup> Portions of this paper were presented at the meetings to the Michigan Academy of Science, Arts and Letters, March 24, 1961. Material included in this report stems in part from a project sponsored by the Personnel Laboratory, Wright Air Development Center, Air Research and Development Command, Lackland Air Force Base, Texas.

### DEVIATION HYPOTHESIS AND ITS ATTENDANT NOTIONS: AN EXPLICATION AND CRITIQUE

The following analysis of Berg's position is based entirely upon his most recent paper (Berg, 1961) except where otherwise noted. Berg does not, however, seem to have made any major conceptual changes in this latest statement from those previously published. Thus we find on page 335 the Deviation Hypothesis quoted directly from an earlier paper; viz.,

Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality (atypicalness) and thus regarded as symptoms (earmarks or signs) are associated with other deviant response patterns which are in noncritical areas of behavior and which are not regarded as symptoms of personality aberration (nor as symptoms, signs, earmarks) (Berg, 1957, p. 159).

First, some purely semantical matters deserve mention if only because certain seemingly crucial terms employed in this statement are apparently intended quite differently than a casual reader—especially a psychologist—might otherwise expect. For example, despite the pathological connotations ordinarily carried by such terms as "abnormality," "deviant," "personality aberration," and "symptoms" both in the natural language and in technical psychological discourse, no such intension is necessarily here implied. "Not merely the behavior deviations such as schizophrenia are included but any abnormal be-

havior, that is, *ab-normal* in the literal sense of being away from normal" (p. 334).

However, even this latter usage appears to be too restrictive to suit Berg, for in reply to Edwards (1959) he explicitly disavows that "deviant response" necessarily implies a departure from a general normative or majority response. "The important thing is that the responses of the criterion group *differ* (italics mine) from the deviant group at a given level of statistical significance . . ." (p. 338). Nor, apparently, is there any importance to be attached to the seemingly asymmetric relation suggested by the use of "criterion group" in contrast to "deviant group." While Berg acknowledges that the deviant group is always a special segment of the general population and that the criterion group is *usually* people in general or a random sample from the "normal" population, this latter need not be the case; ". . . any group which serves as a basis for comparison may be used for criterion purposes" (p. 342).<sup>2</sup> In this connection also it should be noted that the decision as to whether a test response should be included in a scoring key is apparently to be based on a test of the hypothesis that the two populations have the *same* relative frequency for that response against a *nondirectional* alternative to the null hypothesis.

Thus, by a deviant response Berg apparently means simply *any response category* (or

<sup>2</sup> Berg seems to have created a further and unnecessary confusion by his choice of labels for the two groups—and especially by his use of criterion group. The phrase "criterion group method" has been traditionally applied to test construction procedures where data external to the test responses themselves are used to identify members of the two groups between which differential test responses are then sought. Where the responses of a general or normative population are being contrasted with those of an identifiable subpopulation contained within it, it is not uncommon to refer to the *subpopulation* as the criterion group—a convention which Berg has completely reversed. Where comparisons are to be made between mutually exclusive subpopulations, external, differential criteria are clearly required to denote or identify members of *each* group. In this latter case it would seem desirable to not confuse the situation by alluding to one group by a term properly applicable by virtue of the method employed to either member of the comparison.

*class*) for which the relative frequency of occurrence in one group differs from that in some other group being considered, and nothing more.

One final semantic clarification is called for—that which stems from the implicit differentiation of "critical" and "noncritical" areas of behavior in the statement of the Deviation Hypothesis. By critical (or "significant") areas of behavior are meant those particular qualities which serve to determine membership in a given class (the criterion or the deviant groups) whereas noncritical behaviors are all those which, while (presumably) also containing valid differential indicators for the class, are not so conceived nor ordinarily used by the community for this purpose. The basis for distinguishing between the two kinds of behaviors is apparently partly a function of the social importance attached to each. "Putting it another way, response patterns in critical areas of behavior have *value* attached to them in a particular culture and they usually bear classificatory labels" (p. 339). In addition to the cultural importance criterion (usually with an attendant verbal labeling of the class), it is apparently also required that all, or at least many, of the critical behaviors be jointly present in all members of the class. For example, "criminals," "delinquents," "executives," and "supervisors" are all given as examples of culturally valued and labeled classes but ones in which behavioral heterogeneity of the members is so large that such categories could not serve as either criterion or deviant groups.<sup>3</sup>

<sup>3</sup> Berg (1961) does not actually give explicit definitions of either critical or noncritical as areas of behavior. The above is an attempt at a somewhat more precise statement of the criteria by which the two kinds of behavior are distinguished as I understand it from the informal discussion and examples Berg does present. One gets the uncomfortable feeling, however, that what really serves to define "valid behavioral categories" in particular cases is not a prior application of both or either of these two criteria. Rather, what seems to be crucial is whether or not some given aggregate of persons, however related and labeled, can be separated from some other group by an empirically constructed scoring key for responses to some set of superficially irrelevant (noncritical) stimuli. If this be so, then the Deviation Hypothesis, whatever its other virtues or faults, becomes very tightly circular indeed.



Let us now consider some of the attendant notions or consequences of the Deviation Hypothesis which Berg (1961) presents (pp. 336–337, 365–372). First, he argues, that one of the consequences is “A definition of deviant responses in simple and operational statistical terms—no hunches, clinical intuition, judgments, ratings, or other subjective appraisals are involved” (p. 336). This statement must certainly pertain primarily, if not exclusively, to deviant responses to potential test stimuli, i.e., to those in the noncritical behavior area. As just one example, in the study by Barnes (1955) which Berg relies on heavily to make his case, judgments or subjective appraisals were rather intrinsically involved in the so-called critical behavior area. The fact that a patient has a recorded diagnosis of schizophrenia or character disorder or something else which can be objectively *transcribed* in no sense make this designation independent of the basis on which the original diagnosis was made—a very highly subjective and judgmental process in most psychiatric settings.

A second proposition is presented as

A deduction [sic!] drawn from a vast array of evidence that virtually any relatively unstructured stimulus content can be used to elicit deviant responses. This is not, of course, central to the hypothesis proper. . . . This is not saying, it should be emphasized, that any item is as good as any other for a particular purpose. Such an assertion would be absurd. But it is saying that for a given verbal item, for example, one would [could?] find a design or sound that would do as well. At least all of the evidence we have thus far garnered forces us to that conclusion (p. 336).

This particular derivative of the Deviation Hypothesis, whether taken as a “deduction” from the statement that “Deviant response patterns tend to be *general* (italics mine)” or whether considered as an *inductive inference* based on the past empirical findings by Berg and his associates, is the primary focus of the present paper.

It is somewhat difficult to see why Berg feels that this proposition is not central to the Hypothesis proper. Assuming that the Deviation Hypothesis says anything at all which is nontrivial and empirically meaningful (i.e., assuming that the definitions of critical and noncritical behaviors can be given independently of the defining character-

istics of the criterion and deviant groups in some such manner as that suggested above), then it would seem that the Deviation Hypothesis asserts just this proposition—that deviant response patterns tend to be *general*, i.e., that they occur in both the critical and noncritical areas of behavior. It is true that the Hypothesis proper does not assert anything about the relative *degree* of separability that can be achieved between the contrasted groups by the various possible classes of test stimuli. In this regard at least, the statement that one could find a design or sound that *would do as well as* would a verbal, content relevant item does assert something beyond what is expressed in the Hypothesis itself. It could be argued, however, that the Deviation Hypothesis has scientific (or better, technological) importance only to the degree that behavior in the noncritical areas can be used to effect a relatively substantial separation of the contrasted groups as compared to that obtained by use of more highly content relevant stimuli. A somewhat similar analysis of the asserted generality of deviant response patterns and some empirical findings pertinent to this issue have been presented earlier by Sechrest and Jackson (1962).

The third implication of the Deviation Hypothesis which Berg presents first reasserts the generality or pervasiveness of deviant responses but then adds another point—that members of any deviant group have in common response patterns in both the critical and noncritical areas. “Thus, an engineer, a schizophrenic, or a genius should each reveal atypical responses which form a pattern peculiar to his special class. The pattern would be largely shared by other members of the same class” (pp. 336–337). The fourth point mentioned is that the Hypothesis offers a conceptual framework which suggests how a variety of problems, not just those of psychopathology, may be attacked. He asserts that the method is applicable to “any class of behavior which is identifiable by valid external criteria” (p. 337). And finally, Berg contends that the Hypothesis serves to explain a wide variety of findings of a seemingly diverse sort and thus provides a unifying principle which *accounts for* such re-

sults—more of this later. Let us turn now to a brief review of some of the empirical findings which Berg cites in support of his position.

#### SOME PRIOR EMPIRICAL RESULTS IN SUPPORT OF THE DEVIATION HYPOTHESIS

Berg reviews a fairly large number of empirical investigations, the results from which he interprets as support for his position. Most of the studies employed the criterion group method and for the most part the findings are based on responses to Berg's Perceptual Reaction Test (PRT). Some findings are reported, however, based on auditory stimuli or on more conventional items such as self-descriptive adjectives, self-report statements, and occupational preferences. Most of these latter findings arise from studies which also employed designs other than the criterion group method. A brief survey indicates that groups differing in psychiatric diagnosis, in chronological age, in degree of mental retardation, in amount of emotional disturbance, and in severity and tenure of physical illness all have been found to differ in their responses to the PRT items. While many of the findings are reported only in terms of a value for some test statistic or the corresponding Type I error probability, in some cases enough data are reported either in Berg's paper or in the original sources cited by him to make some estimate of the degree or magnitude of the relationships or differences involved.

By all means the most dramatic support comes from the study by Barnes (1955) mentioned earlier. In this study PRT responses were obtained from a heterogeneous group of 1,700 normals and from 546 patients of mixed psychiatric diagnoses. Males and females were treated separately. From the mixed abnormals a subset of psychotics was identified within each sex group and from among these, smaller groups of male and female schizophrenics were also isolated. In addition, a group of 120 male character disorders were also identified and were used separately in some of the analyses. Barnes partitioned each of the identifiable groups of subjects into two parts. Then by means of a double cross-validation design he developed

empirical scales for contrasted groups on each pair of subsamples separately and estimated their validities on the remaining cases. The results reported indicate clearly that for all scales developed, the contrasted groups were well separated. Even for the one truly *differential* diagnostic key—that contrasting male psychotics with male character disorders—the data reported indicate that by use of an optimal cut score it is unlikely that the false positive or false negative rates would be much higher than 25% to 30%, assuming comparable base rates. Considering the brevity of the PRT, the apparent irrelevance of its content to psychiatric nosology, and the unreliability and often fortuitous nature of differential diagnostic classification, this latter finding, at least, is indeed impressive. If subsequent studies should indicate that equally discriminating scales can be built to separate, say, neurotics from psychotics, or process from reactive schizophrenics, or neurological organics from functional psychotics, then indeed the PRT and the Deviation Hypothesis will have to be reckoned with by diagnosticians with other pet instruments and methods.

The study to be reported below, however, does not pursue this aspect of the problem. It is concerned with differential classification, but within the domain of nonpathological personality organization.

#### INVESTIGATION OF THE RELATIVE IMPORTANCE OF TEST ITEM CONTENT FOR THE ASSESSMENT OF NONPATHOLOGICAL PERSONALITY ATTRIBUTES

The critical behavior domain employed in this investigation derives from peer ratings among close associates. An extended series of investigations reviewed elsewhere by the writer (Norman, 1963b) which employed scales drawn from Cattell's reduced personality sphere set have each demonstrated the existence and relative orthogonality of five personality factors. The populations on which these results are based range from Air Force personnel, through college undergraduates to applicants for graduate school programs in clinical psychology. The five factors have tentatively been designated (each by the name of one of its two poles)



Surgency (or Extroversion), Agreeableness, Dependability, Emotional Stability, and Culture. Details of the structure, comparabilities across samples, salient scales for each factor, etc. are presented and discussed in the review previously mentioned (Norman, 1963b). Suffice it to say here that each of these five factors is quite clearly defined and easily interpreted in terms of polar clusters of highly homogeneous behavioral descriptors and that this structure was found in each of the samples described below.

The present study was intended to compare the relative effectiveness of responses to content irrelevant stimuli with those made to more highly relevant test items in terms of their ability to predict the status of individuals on each of the five peer rating factors. In the first analysis three kinds of test stimuli were employed: personality descriptive adjectives, occupational titles, and figural drawings. The adjectives, constituting a set of maximally content relevant stimuli, were drawn from a large pool of trait descriptive terms on the basis of judged high relevance to one or another of the five factors. The terms so chosen were then paired such that each member of a pair was relevant to a different factor but was matched with the other term in the pair on a particular rated desirability stereotype. The 200 forced-choice, self-report items constructed in this manner were typed in booklet form and called the Descriptive Adjective Inventory (DAI).

A set of occupational titles was chosen as an intermediately relevant set of stimuli. Judgments of the factor relevance of preference responses to each of these stimuli proved to be very difficult to make and the judges expressed little confidence in their ability to do this. Again ratings of the desirability of expressed preferences for each occupation were also obtained and a set of 60 binary forced-choice items were constructed in the manner described above. These items were then typed in booklet form and named the Occupational Preference Inventory (OPI).

Finally, the set of content-irrelevant stimuli used were the 400 figural designs which are contained in the Research Edition of the Welsh Figure Preference Test

(WFPT). In this test, the respondent simply records whether he "likes" or "dislikes" each of the designs.

The samples of subjects employed were paid male volunteers from a number of fraternities, residence halls, and one cooperative housing unit at the University of Michigan. Sample I was composed of 82 men, mostly seniors, from whom the peer nomination ratings and test performances were collected in the Spring semester of 1960. Sample II consisted of 215 fraternity men, mostly sophomores and juniors, tested in the Fall semester of 1960. Sample III (used together with Sample II in a second analysis reported below) was composed of 241 residence hall men, primarily freshmen, sophomores, and juniors, tested in the Spring semester of 1961.

Contrasted criterion groups on each factor were formed in Sample I by splitting each distribution of factor scores (derived from the peer ratings) at its median. Percentages of each subgroup on each factor responding to each item alternative on each of the tests were computed. From these values percentage difference discrimination indices (item validities) and percentage endorsement indices were computed. A set of five preliminary empirical scoring keys was constructed for each of the three inventories by simply including all responses in a given key for which the item validity indices were greater than a selected minimal value.<sup>4</sup>

The answer sheets for each subject in Sample II were then scored on the sets of preliminary empirical keys just described. The correlations among these scales and the criterion factor ratings are given in Table 1. The values in boldface and brackets are

<sup>4</sup> This minimal value corresponded approximately to a 5% significance level. However, since a *measure of validity* was used instead of a *significance test statistic* for such a measure, some responses for items with nearly even response splits probably had significance levels which exceeded .05 and some, especially those with extreme endorsement rates, certainly had Type I error likelihoods well below .01, even for these small samples. The use of a *measure of validity* (instead of *merely* a test of significance) for purposes of selecting keyable response categories for empirical scales is a practice which can be strongly defended for reasons both too numerous and too complex to be expanded upon here.

TABLE 1

CORRELATIONS AMONG THE RATING FACTOR SCORES AND SCORES ON THE PRELIMINARY EMPIRICAL KEYS FOR THE DAI, OPI, AND WFPT

		Ratings					DAI					OPI					WFPT				
		I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
Ratings	I																				
	II	08																			
	III	-27	48																		
	IV	20	55	10																	
	V	02	37	63	18																
DAI	I	[40]	-23	-30	-08	-11															
	II	-21	[25]	22	02	03	-31														
	III	-33	06	[40]	-27	11	-46	33													
	IV	22	-04	-23	[24]	-09	16	-19	-65												
	V	11	-07	-21	13	[16]	18	-43	-63	36											
OPI	I	[03]	06	-01	-01	15	(23)	-02	-06	-09	09										
	II	-04	[07]	09	-00	03	01	(07)	-11	05	08	-17									
	III	07	04	[-08]	-03	-26	-14	05	(16)	-04	-27	-32	-25								
	IV	-12	01	04	[-05]	-07	-27	-05	07	(04)	03	-75	13	24							
	V	03	00	05	01	[29]	18	-11	-25	04	(35)	44	29	-68	-34						
WFPT	I	[00]	07	10	-04	12	(03)	11	09	-09	-10	(03)	-16	-02	-02	03					
	II	03	[08]	08	04	15	04	(09)	-01	01	00	-03	(-07)	-07	09	10	75				
	III	-01	-14	[-22]	06	-23	-02	-07	(-06)	14	08	-05	11	(-12)	09	03	-22	-05			
	IV	-07	04	02	[-03]	-01	-13	11	05	(-04)	-16	03	-09	08	(-06)	-09	-06	-24	-40		
	V	01	-07	-10	07	[-04]	01	-14	-11	11	(19)	-07	24	-11	08	(09)	-50	-31	52	-34	

Note.—Based on Sample II (total  $N = 215$ ). Decimal points omitted.



cross-validation estimates of validity of the scales against the peer nomination rating factor scores. The values in boldface and parentheses are the correlations between instruments for corresponding factor keys.

The first thing to note is that the keys for the WFPT have essentially zero validities. The largest magnitude, that for the Factor III key, is in fact *negative*. The validities for the OPI keys are likewise essentially zero with the one exception of the .29 value for Factor V. By way of contrast all five of the validity coefficients for the DAI keys are positive and statistically significant using a .05 Type I error level.

Taken together, these results appear to argue rather forcibly against Berg's contention that test item content is unimportant in the construction of valid empirical keys. Nor, I believe, can these findings be dismissed as irrelevant because of some defect in the design or procedures employed. If it is argued that the size of Sample I is too small to be used to develop criterion group empirical keys, or that the use of median splits on the rating factor score distributions misclassifies criterion cases, then how does one explain the significant and, in some cases, appreciable validities obtained for the DAI scales? If the relatively small number of available response alternatives on the OPI is mentioned, it must be pointed out that the WFPT has 800 potentially keyable response categories whereas the DAI has only half that number. Thirdly, if it is suggested that perhaps the peer nominations do not yield valid, "operationally clean" criteria for identifying the contrasted groups, one must point out that the factor structure based on these ratings is exceptionally clear-cut and that the sole basis for classifying subjects into groups is on the basis of explicitly labeled (and undoubtedly socially valued) behavioral homogeneity as judged by groups of intimate associates. And finally, the point made by Berg that the stimuli must not be too highly structured so that response sets are free to operate has certainly been met by all three devices and especially by the WFPT whose stimuli are at least as unstructured as those on the PRT.

Although the analysis just discussed was

hardly encouraging with regard to the potential value of further work with empirically keyed figural preference responses for the measurement of the five rating factors, there were several reasons to pursue this approach at least a bit further. Berg and his associates had, after all, generated a rather large number of apparently successful demonstrations of this approach—to be sure with other groups, and in other kinds of settings, but with similar stimuli. A second and more immediate reason was the nature of the research program in which we were engaged. The general problem under attack was to develop individual performance measures of these five factors that would be suitable for use in a personnel selection context where the peer nomination method was administratively unfeasible. Because applicants for the training program for which these tests were to be used as selectors are highly motivated to be accepted, some assurance was needed that applicants would not be able to fake their responses to gain admission. Since the peer nomination scales defining the factors each have a clearly discernible positively valued pole (in this setting, at least), simple self-ratings on these scales or self-reports on typical inventory scales were expected to be highly vulnerable to any attempts to dissimulate responses. One possible approach to this problem, of course, would be to employ test stimuli which bear no apparent relation to the variables to be measured.<sup>5</sup> If such stimuli could be used to develop scoring keys with substantial validities, then the problem presented by a set to "fake good" in a respondent would be greatly diminished. Accordingly a second analysis was done.

The 456 subjects from Samples II and III were divided by peer rating groups into a pair of cross-validation samples of 228 men each—hereafter referred to as Samples A and B. These two samples were matched as

<sup>5</sup> A quite different approach to this problem and one which permits the use of content relevant test stimuli, has been described elsewhere (Norman, 1963a). The evidence so far available indicates that the method presented in that paper holds considerable promise for an effective solution of the problem of faking in at least certain kinds of assessment contexts.

closely as possible in terms of residence (fraternity or dormitory), academic class standing, and similarity of their criterion factor structures based on the peer rating data. Each sample was then split trichotomously on each criterion factor with about 30% of the sample in each of the extreme classes and the remaining cases in the middle group. For each item response category on the WFPT and on the DAI (as well as on two other inventories which need not concern us here), overall endorsement indices and the five matrices of percentage difference discrimination indices among the trichotomous groups—one for each of the criterion variables—were computed. This was done separately for each sample and for each test.<sup>6</sup>

The discrimination indices (between the two extreme criterion subgroups) for the DAI items as calculated from the Sample A data were plotted against those obtained from Sample B. An inspection of the scatter plots indicated the relationship to be moderately high and positive for each of the five criterion factors. In addition, appreciable numbers of items had discrimination indices exceeding 15% in magnitude in both samples and in some cases they extended as high as 30% to 35%. For this test, items for which a discrimination index of 15% or higher on one sample reversed in sign on the other sample were extremely rare. Quite clearly, these data could be (and were) used to build empirical scoring keys for which validities on all five criteria would be maintained in cross-validation. The results of that effort have been reported elsewhere (Norman, 1963a) and will not be repeated here.

However, when the corresponding analysis was done for the items of the WFPT, the results were completely discouraging! Not only were the discrimination indices smaller than for the content relevant items in the other inventories analyzed, it was extremely rare to find one as large as 15% between the

high and low criterion groups. What is worse, such items as were found with validities in this range on one sample were as likely as not to show a drop in validity to zero or even a reversal of sign in the other sample. The scatter plots of the item discrimination indices between the two samples were, in fact, almost perfectly circular and the marginal distributions looked, for all the world, like random error distributions. It was perfectly clear that any attempt to construct scoring keys from such data in the hope of being able to cross-validate them was a fool's enterprise and consequently the effort to utilize such test materials for our purposes was abandoned at this point.

#### DISCUSSION OF RESULTS AND THE QUESTION OF THE THEORETICAL SIGNIFICANCE OF THE DEVIATION HYPOTHESIS

Whatever other implications may be drawn from these findings, one thing certainly is clear—the term “general” in Berg's formulation of the Deviation Hypothesis cannot be taken in anything like the unqualified sense in which he has used it. And in at least certain kinds of personality assessment contexts, the use of relevant test item content is of rather considerable importance.

But if one grants these points (as I believe he must in the face of the data presented above), the far more intriguing questions only briefly alluded to earlier follow at once. In what particular kinds of assessment contexts, for which classes of subjects, and with what degree of success can which specific kinds of noncritical behaviors be employed to measure personality or interests or any other critical attributes or status variables? For *purely pragmatic purposes*, such detailed knowledge is of the essence if we are ever to emerge from the “let's-try-everything-and-see-what-works-this-time” approach to test development. In this connection, the extensive bibliography which Berg and his associates (Berg, 1961) have generated is bound to be useful—especially as it continues to accumulate and is reported in increasing detail.

But in a wider sense is this sufficient for anything but the most immediate applied

<sup>6</sup> Each subject actually took each test twice: first under standard instructions to be as honest and accurate as possible and secondly under instructions to fake the most desirable responses possible to gain admission to the training program referred to above. We will be concerned here only with the analysis of the data obtained under the first set of instructions.



or technological purposes? Does even a very extensive compendium of such findings—even when subsumed under some such statement as that presented in the Deviation Hypothesis—constitute an adequate theoretical structure for an empirical science? In my opinion the answer is, No! Even assuming that enough data are eventually forthcoming so that we will know under what conditions, for whom, and with what degree of success, which stimuli can be depended upon to yield valid assessments of which behavioral phenomena—at that point we will have our “engineer’s handbook” and little else. What is more crucial in my opinion is that by any consistent adherence to the distinction implied by the conceptions of critical and non-critical areas of behavior, we will have explicitly precluded the development or construction of *any theoretical system* capable of yielding an adequate “explanation” of behavioral phenomena. What is the distinction, after all, between the terms critical and noncritical when applied to areas of behavior but an admission of the incompleteness of our present knowledge? An adequate empirical theory, from the present viewpoint at least, must be more than a mere catalogue of what goes with what. Nor does it seem particularly fruitful to codify and perpetuate, in the major postulate of a formulation proposed for this purpose, a distinction that simply reflects our current naivete and social values. On these grounds, it seems to me that whatever pragmatic virtue the Deviation Hypothesis might have for goading

researchers to explore new possibilities for useful test stimuli, its purported theoretical value *to account for or explain* specific behavioral phenomena or relationships is, at the very best, illusory.

#### REFERENCES

- BARNES, E. H. The relation of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, 51, 286-290.
- BERG, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, 40, 61-71.
- BERG, I. A. Deviant responses and deviant people: The formulation of the deviation hypothesis. *J. counsel. Psychol.*, 1957, 4, 154-161.
- BERG, I. A. The unimportance of test item content. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality assessment*. New York: Van Nostrand, 1959. Pp. 83-99.
- BERG, I. A. Measuring deviant behavior by means of deviant response sets. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 328-379.
- EDWARDS, A. L. Social desirability and personality test construction. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality assessment*. New York: Van Nostrand, 1959. Pp. 100-118.
- NORMAN, W. T. Personality measurement, faking, and detection: An assessment method for use in personnel selection. *J. appl. Psychol.*, 1963, 47, in press. (a)
- NORMAN, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *J. abnorm. soc. Psychol.*, 1963, in press. (b)
- SECHREST, L., & JACKSON, D. N. The generality of deviant response tendencies. *J. consult. Psychol.*, 1962, 26, 395-401.

(Received August 31, 1962)

## CLIENT LIKABILITY: A VARIABLE IN THE STUDY OF PSYCHOTHERAPY<sup>1</sup>

NORTON STOLER

*Wisconsin Psychiatric Institute, University of Wisconsin*

A study designed to investigate the rateableness of client likability (CL), and the relationship of CL to success in psychotherapy. 10 clients were rated as to their likableness by 10 raters from 2 recorded segments taken from the therapy of each of the clients. The clients were classified into more successful and less successful categories prior to this study. 2 levels of rater familiarity with the cases were employed. It was found that (a) CL was rated reliably by the familiar raters beyond the .01 level, (b) the more successful clients were liked significantly greater than the less successful clients, beyond the .05 level, and (c) CL correlated significantly with the Experiencing Strand of Rogers' Process Scale. It was concluded that CL can be reliably rated and that success in therapy is possibly related to this variable.

The background for this study is a very global and personal reaction to clients. Simply stated, this reaction is that for some clients one can feel immediately more compassionate and optimistic, while to others one feels less personally moved and also more discouraged about being helpful. This study attempts to determine the generality of this reaction, by having raters give their own personal liking for clients, as heard in recorded therapy interviews. The purpose is to see if such a personal and subjective reaction can be reliably measured and to investigate the relevance of client likability in psychotherapy.

More operationally stated the research reported here investigates the following:

1. The reliability of a scale to measure client likability.
2. The relationship between the likability ratings of clients and the success of these clients in therapy.
3. The factor of familiarity with the clients by the raters and the likability ratings for these clients.
4. The change in likability from early to later interviews.

5. The relationship between the client's rated likability and a measure of the client's in-therapy behavior as reflected by Rogers' Process Scale (Walker, Rablen, & Rogers, 1960).

### PROCEDURE

Ten raters were given 20, 2-minute, taped segments of therapist and client interactions drawn from 10 recorded therapy cases.<sup>2</sup> Prior to this study five of the cases were classified as more successful and five as less successful, on the criteria of therapist ratings of outcome, patient ratings of outcome, and a self-concept *Q* sort. Two segments were used from each case, one from an early interview and one from a late interview. The 20 segments were presented to each judge in a random order; this was made possible by having each 2-minute segment on a separate tape spool.<sup>3</sup> The raters were then given the following instructions with the 20 segments.

### *Instructions for the Rating of Client Likability*

After listening to the taped segment of the client and therapist interchanges, make a rating on the continuum from liking to disliking of the client.

You may make a mark at any place along the scale, you are not confined to the points that are numbered.

<sup>1</sup> Partial support for this research was received from the National Institute of Mental Health, Grant No. M3496.

This is an extended version of a paper presented at the national convention of the American Psychological Association in New York City, September 1961.

<sup>2</sup> The 10 clients rated in this study were from 10 fully recorded cases from the University of Chicago Counseling Center.

<sup>3</sup> These 20 segments were drawn from a large pool of segments of these 10 recorded cases used by Tomlinson and Hart (1962) in a validation study of the Process Scale.



Scale point number one (1) is for a positive liking reaction to the client, while a check mark at number six (6) would be a disliking reaction to this client. Marking any place along the scale between these two points will represent the magnitude of your liking or disliking, depending on the closeness to the end of the scale.

Often it is our experience that we have feelings and reactions to people, but do not necessarily draw our attention to these feelings. This rating task asks of you to *look at the specific liking or disliking feeling that this client brings out in you*, and to mark on the scale the point that best describes *your reaction* along this continuum.

The 10 raters that made their likability judgments were all people with some experience as therapists. Five of these raters had previously made ratings on the Experiencing Scale (one of the scales in the Process Scale) on the same clients and had heard many segments from the recorded interviews. The other 5 raters were hearing these clients for the first time. The distinction will be made between these two groups of raters as the familiar and unfamiliar raters, respectively.

A summary of the design of this study is as follows: there are 20 segments drawn from 10 therapy cases, on several criteria prior to this study, 5 of the cases were classified as more successful and 5 as less successful. From each case there is a segment from an early interview and one from a late interview. Of the 10 raters, 5 were familiar with the clients from previous process rating and 5 were unfamiliar with these clients.

## RESULTS

Table 1 presents the summary of the analysis of variance of the likability ratings.<sup>4</sup> It was found that the more successful clients received a significantly higher mean likability rating than the less successful clients (2.6 versus 3.5). In addition, it was found that the five familiar raters gave significantly higher likability ratings than the five unfamiliar raters (2.8 versus 3.3).<sup>5</sup>

No significant change in likability was found from early to late therapy interviews. The more successful clients were more liked early in therapy by raters and they maintained this advantage over the less successful clients.

The results of rater reliability of this subjective rating task are indicated in Table 2. Table 2, below the diagonal, presents the interrater reliability for the familiar raters which indicates that the raters who were familiar with the clients showed good agree-

<sup>4</sup> The author wishes to acknowledge A. Alexander of the Wisconsin Psychiatric Institute for his consultation on problems of design and analysis.

<sup>5</sup> The Likability Scale ranges from 1 to 6 with liking at 1 and disliking at 6.

TABLE 1  
SUMMARY OF THE ANALYSIS OF VARIANCE OF THE LIKABILITY RATINGS

Source	df	SS	MS	F
Group (A) (More successful-less successful)	1	36.98	36.98/a	8.56*
Time (B) (Early interviews-late interviews)	1	.41	.41/c	—
Rater familiarity (C)	1	16.25	16.25/b	6.09*
A × B	1	.08	.08/c	—
A × C	1	2.42	2.42/c	—
B × C	1	.40	.40/c	—
A × B × C	1	.72	.72/c	—
Error terms				
a Subject's within groups (More successful-less successful)	8	34.60	(a) 4.32 (a)/c	—
b Raters within familiarity	8	21.39	(b) 2.67 (b)/c	—
c Residual	176	1,390.45	(c) 7.90 (c)	—
Total (N - 1)	199	1,503.70		

\*  $p < .05$ .

TABLE 2

INTRATER RELIABILITY OF FAMILIAR AND UNFAMILIAR RATERS ON THE LIKABILITY SCALE

	F	G	H	I	J	
Familiar raters	A	.08	.36	.55**	.17	F
	B	.52**	.19	.04	.26	G
	C	.47*	.53**	.60**	.33	H
	D	.73**	.70**	.42*	.40*	I
	E	.75**	.36	.54**	.73**	J
	A	B	C	D	E	

Note.— $df = 18$ .\* $p < .05$ .\*\* $p < .01$ .

ment as to whom they liked and whom they disliked. The average rater reliability for this group of familiar raters was .58, which is beyond the .005 level.

In contrast, the unfamiliar raters were unable to show this degree of reliability. Table 2, above the diagonal, presents the matrix of rater reliability coefficients found among the unfamiliar raters. The average rater reliability for this group is .30 which is significantly lower than the average reliability of the familiar raters.

The familiar raters had made Experiencing Scale ratings on these clients several months prior to the time of the likability ratings. This made it possible to determine the relationship between client likability and this process scale variable. Before presenting the results of the correlations between these two

scales a word of introduction to the Experiencing Scale.

The concept of experiencing is one of the central notions in Rogers' Process Scale. The details of this concept can be found in Gendlin and Tomlinson (1962); Gendlin (1961) Walker, Rablen, and Rogers (1960), and Tomlinson and Hart (1962). In brief, the Experiencing Scale deals with the degree to which the client focuses on his personal experiencing and this ranges from "remoteness from experiencing and unawareness" at the low end of the scale with "decreasing remoteness and increasing awareness" at the middle range and "lives in process of experiencing and uses it as a major referent" at the upper end of the scale (Gendlin & Tomlinson, 1962).

The five familiar raters who made the Experiencing ratings and the Likability ratings had a significant correlation between these two scales. That is, the more the client is liked the higher the process rating and the more he is disliked the lower the process rating. Table 3 presents these correlations which ranged from  $-.37$  to  $-.70$ .

TABLE 3

CORRELATIONS FOR THE FAMILIAR RATERS BETWEEN THE LIKABILITY AND THE EXPERIENCING RATINGS

Familiar raters	Correlation coefficients
A	$-.63^{**}$
B	$-.61^{**}$
C	$-.37^{*}$
D	$-.51^{*}$
E	$-.70^{**}$

Note.— $df = 18$ .\* $p < .05$ .\*\* $p < .01$ .

## DISCUSSION

The problem of scaling something so variable and subjective as the degree of liking of a client just from a taped segment of the interview raises the question of reliability. The average rater reliability of .58 for the



familiar raters indicates that some consistent cues are being used in making these judgments. This degree of rater reliability indicates that client likability may be a general characteristic on which raters can agree, that is, once they have become more familiar with the client.

The finding that the familiar raters liked the clients significantly more than the unfamiliar raters may indicate that raters change in their attitude toward the clients as they hear more of their therapy.

In addition, since the raters have all had experience as therapists and are familiar with a variety of client behavior, their judgment of liking may depend on what they consider to be "good" in-therapy behavior. This interpretation is supported by the significant correlations found between the client's rated likability and his rated level of process on the Experiencing Scale. These findings point to a possible source of biasing in the rating of psychotherapy material.

The clients whose experience of psychotherapy is less successful may be the people who are much more difficult to like and it is this behavioral characteristic of degree of likability that may be related to the less successful outcome of therapy. Rogers (1957) has presented the theoretical formulation that if certain therapeutic conditions are provided by the therapist, and in turn if these conditions are *perceived* by the client, then constructive personality change will occur.

It is possible that the therapist may be unable to provide effective therapy to a

client he cannot like and the client who has had little experience in life in being liked may be unable to perceive that the therapist really cares for him. Future concentration on the phenomenon of Client Likability as an interactive factor in the relationship between the therapist and the client may account for some of the variance that contributes to the effectiveness of psychotherapy.

In conclusion, this study raises two questions for research in psychotherapy, one, whether the variable of Client Likability is a "halo effect" that contaminates the validity of outcome criteria and process measures, and two, whether Client Likability is a relevant personality variable which is a causal factor in the success of psychotherapy.

#### REFERENCES

- GENDLIN, E. T. Experiencing: A variable in the process of therapeutic change. *Amer. J. Psychother.*, 1961, 15, 233-245.
- GENDLIN, E. T. *Experiencing and the creation of meaning*. Glencoe, Ill.: Free Press, 1962.
- GENDLIN, E. T., & TOMLINSON, T. M. Experiencing scale. Madison: University of Wisconsin, Wisconsin Psychiatric Institute, 1962. (Mimeo)
- ROGERS, C. R. The necessary and sufficient conditions of therapeutic personality change. *J. consult. Psychol.*, 1957, 21, 95-103.
- TOMLINSON, T. M., & HART, J. T. A validation study of the process scale. *J. consult. Psychol.*, 1962, 26, 74-78.
- WALKER, A., RABLEN, R. A., & ROGERS, C. R. Development of a scale to measure process changes in psychotherapy. *J. clin. Psychol.*, 1960, 16, 79-85.

(Received September 19, 1962)

## NOTES AND COMMENTS

### PSYCHOTICISM OR TEN PSYCHOTIC SYNDROMES?

H. J. EYSENCK

University of London

In their recent article, Lorr, McNair, Klett, and Lasky (1962) report a factor analysis of a rating schedule for 296 psychotic inpatients, concluding that "ten postulated psychotic syndromes were confirmed, in varying degrees" (p. 189). They also declare that "there is no evidence in support of Eysenck's contention that there is a single second-order factor of psychoticism" (p. 189). As neither conclusion is justified by their data, a brief examination of their failure to use the technique of factor analysis properly may be in order.

It is universally agreed that to obtain evidence for a given factor, this factor must be represented adequately (a) in the tests used, and (b) in the population studied. It is clear that no second-order factor of general intelligence will be found if all the subjects tested have very similar IQs; where there is no variance there can be no covariance! The proper test of the existence of a factor of psychoticism requires a population in which some have high scores, others low scores on this factor; the hypothesis cannot be tested on a population in which by definition every member is high (probably almost uniformly high) on this factor. Clearly no correlations can be generated under these conditions, which are optimal for discovering group factors, but which do not permit of a proper examination of the existence of a general factor of psychoticism. In such studies as have disclosed this factor, psychotic, neurotic, and (usually) normal subjects have always been included (Eysenck, 1952a, 1952b, 1955, 1960; Trouton & Maxwell, 1956), and more adequate sampling than that provided by Lorr et al. (1962) is a prerequisite if any acceptable conclusions are to be derived about the existence or otherwise of a factor of psychoticism.

The tests used to define this factor have usually been of an objective, performance type, such as reminiscence in pursuit-rotor learning (Eysenck, 1961), mental speed (Eysenck, 1953), visual acuity and accommodation (Eysenck, Granger, & Brengelmann, 1955), pegboard placing (Eysenck, 1952b), object sorting, performance speed and overinclusiveness tests

(Payne & Hewlett, 1960), and so forth; they are characterized by the fact that they give scores on which there is *continuity* between normal, neurotic and psychotic subjects. Lorr et al. used ratings on the majority of which a positive entry would suffice to classify the subject as grossly abnormal; such ratings are inadequate to dis close the existence of a factor of psychoticism should it exist, which is based on the hypothesis of continuity (Eysenck, 1952a). In both respects (tests used and population studied) the experiment under consideration fails to be relevant to the conclusion reached; it was not designed to investigate Eysenck's hypothesis, and it did not enable any conclusions to be reached regarding it.

Neither does the discovery of 10 "syndromes" appear to be a justifiable conclusion from the data reported. Most of these factors seem to be, not syndromes validated, but symptoms shown to be rated reliably. Tests and items used for a factor analysis should be logically independent; there is no such independence in the items here used. Consider the following items which not unnaturally hang together in one factor: "Speaks to voices"; "Reports voices accuse or blame him"; "Reports voices tell him to do things"; "Reports voices threaten him." While not entirely identical, these items are so similar that they cannot be regarded as anything but tautologous; correlations between them have the status of reliability coefficients. It is not advisable to include such reliability coefficients in what purports to be a factor analysis devoted to the discovery of "syndromes." It is possible that similar syndromes would still be discovered even after the removal of repetitive items of this kind, which abound in the scale, but it is impossible to arrive at any secure judgment on the basis of the report. We must infer that the conclusions drawn by Lorr et al. are not justified by their own data.

#### REFERENCES

- EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality: II. Experimental. *J. Pers.*, 1952, 20, 345-384. (a)



- EYSENCK, H. J. *The scientific study of personality*. London: Routledge & Kegan Paul, 1952. (b)
- EYSENCK, H. J. La rapidité du fonctionnement mental comme mesure de l'anomalie mentale. *Rev. Psychol. appl.*, 1953, 4, 367-377.
- EYSENCK, H. J. Psychiatric diagnosis as a psychological and statistical problem. *Psychol. Rep.*, 1955, 1, 3-17.
- EYSENCK, H. J. (Ed.). *Handbook of abnormal psychology*. New York: Basic Books, 1960.
- EYSENCK, H. J. Psychosis, drive and inhibition: A theoretical and experimental account. *Amer. J. Psychiat.*, 1961, 118, 198-204.
- EYSENCK, H. J., GRANGER, G. W., & BRENGELMANN, J. C. *Perceptual processes and mental illness*. New York: Basic Books, 1955.
- LORR, M., MCNAIR, D. M., KLETT, C. G., & LASKY, J. L. Evidence of ten psychotic syndromes. *J. consult. Psychol.*, 1962, 26, 185-189.
- PAYNE, R. V., & HEWLETT, J. H. G. Thought disorder in psychotic patients. In H. J. Eysenck (Ed.), *Experiments in personality*. Vol. 2. London: Routledge & Kegan Paul, 1960.
- TROUTON, D. S., & MAXWELL, A. E. The relation between neurosis and psychosis. *J. ment. Sci.*, 1956, 102, 1-21.

(Received June 25, 1962)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 2, 180-181

## CANONICAL VARIATES AND SECOND-ORDER FACTORS:

### A REPLY

MAURICE LORR, DOUGLAS M. MCNAIR, C. JAMES KLETT, AND JULIAN J. LASKY<sup>1</sup>  
*Veterans Administration Hospital, Washington, D. C.*

Eysenck argues that our statement, "There is no evidence in support of Eysenck's contention that there is a single second-order factor of psychoticism," is unjustified. His arguments are for the most part based on a confusion or blurring of the difference between (a) second-order factors in the Tucker-Thurstonian sense (Thurstone, 1947, p. 411), which are factors obtained from the correlations among the first-order factors, and (b) canonical variates. The canonical variate procedure is a way of determining the minimum number of dimensions needed to describe differences between  $c$  groups on  $q$  quantitative variables (Lubin, 1950). Eysenck has utilized canonical variates in virtually all of his demonstrations concerning psychoticism as an independent dimension. Eysenck (1962, p. 50) at times recognizes the differences between the two approaches when he approvingly quotes Slater as saying there is little justification for expecting factor analysis vectors to coincide with those defined by discriminatory analysis of the same data. With this we agree. But in his writings Eysenck repeatedly refers to both types of dimensions as though they were identical (*vide* Eysenck, 1953, p. 53; 1960, p. 15). For example, of Trouton and Maxwell's two first-order factors he says, "These are clearly identifiable as neuroticism and psychoticism,"

implying the presence of second-order factors at the first-order level (Eysenck, 1960, p. 14).

The present authors, in their analysis, demonstrated the presence of three second-order factors among the correlations of the primary syndromes and simply inferred that a single second-order factor was insufficient to account for the inter-correlations. On the basis of this inference it was concluded that Eysenck's assertions concerning psychoticism as a second-order dimension were not tenable. Eysenck's other arguments in support of his between-groups dimensions are simply irrelevant to the issue of the existence of a single common second-order factor. It would thus be well for Eysenck to clarify his thinking or his language in discussing psychoticism in the future.

In passing it can be said that Eysenck has never demonstrated that a single psychoticism factor is sufficient to account for differences between normals, neurotics, and psychotics. He has merely shown that for the particular four or five arbitrary tests administered two canonical variates are sufficient to account for differences between the groups. Introduction of say six other cognitive tests could easily require two additional canonical variates. As Cattell has pointed out, meaningful typing demands a sampling from a specified domain, which Eysenck does not do. The canonical variate dimensions identified by Eysenck are possibly useful in separating groups but tell us very little about the nature of the groups separated.

<sup>1</sup> All authors are also affiliated with Veterans Administration Hospital, Perry Point, Maryland.

Eysenck's second line of argument is broader, more sweeping, intended to sway the reader unfamiliar with test theory or factor method, and to intimidate the enemy. Any novice knows that, because of chance error, a certain degree of redundancy is essential in measurement. In measuring the ability to multiply, items included in a test are to a certain extent tautologous. One perfectly reliable item would be sufficient to differentiate persons at any point along an ability scale. Likewise in psychopathology an excitement syndrome is manifested in hurried speech, rapid body movements, elevated mood, attention seeking, controlling behavior, a loud, boisterous voice, and by an attitude of superiority. Are these elements tautologous? Inspection of the groupings reported in the article (Lorr, McNair, Klett, & Lasky, 1962) and of the original scale (Lorr, Klett, McNair, & Lasky, 1962) should convince any skeptic. The internal consistencies for the scales defining the syndromes range from .75 to .92. Eysenck's assertion would require much higher values. The majority of the scales (42 out of 77) were nine-point continua and not dichotomous as Eysenck erroneously states. It is patent that a dichotomous item does not necessarily require an assumption of discontinuity which is not made by the authors. Eysenck also completely ignores the important fact that the syndromes were postulated in advance and represent the findings from more than a dozen reports by Degan (1952), Guertin (1952), Wittenborn (1951), Lorr, Jenkins, and O'Connor (1955), and Lorr, O'Connor, and Stafford (1957).

In summary, Eysenck has asserted (perhaps carelessly) that (a) a single canonical variate (called psychoticism) is sufficient to separate psychotics from normals and neurotics; (b) there is a single second-order factor common to syndromes descriptive of psychotic symptoms. These two assertions are not identical and have been

confused by Eysenck. Our assertion concerning psychoticism refers only to the latter. The assertion that the syndromes isolated are "invalid," mere tautologous congeries, has no substance. A similar charge could be made, with as little justification, against virtually all factor studies.

## REFERENCES

- DEGAN, J. W. Dimensions of functional psychosis. *Psychometr. Monogr.*, 1952, No. 6, 1-41.
- EYSENCK, H. J. *The structure of human personality*. London, England: Wiley, 1953.
- EYSENCK, H. J. (Ed.) *Handbook of abnormal psychology*. New York: Basic Books, 1960.
- EYSENCK, H. J., & CLARIDGE, G. The position of hysterics and dysthymics in a two-dimensional framework of personality description. *J. abnorm. soc. Psychol.*, 1962, 64, 46-65.
- GUERTIN, W. H. A factor-analytic study of schizophrenic symptoms. *J. consult. Psychol.*, 1952, 16, 308-312.
- LORR, M., JENKINS, R. L., & O'CONNOR, J. P. Factors descriptive of psychopathology and behavior of hospitalized psychotics. *J. abnorm. soc. Psychol.*, 1955, 50, 78-86.
- LORR, M., O'CONNOR, J. P., & STAFFORD, J. W. Confirmation of nine psychotic symptom patterns. *J. clin. Psychol.*, 1957, 13, 252-257.
- LORR, M., KLETT, C. J., MCNAIR, D. M., & LASKY, J. J. *Inpatient Multidimensional Psychiatric Scale*. (Manual) Washington, D. C.: Veterans Administration, 1962.
- LORR, M., MCNAIR, D. M., KLETT, C. J., & LASKY, J. J. Evidence of ten psychotic syndromes. *J. consult. Psychol.*, 1962, 26, 185-189.
- LUBIN, A. Linear and non-linear discriminating functions. *Brit. J. Psychol., Statist. Sect.*, 1950, 3, 90-104.
- THURSTONE, L. L. *Multiple-factor analysis*. Chicago, Ill.: Univer. Chicago Press, 1947.
- WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, 15, 290-302.

(Received September 1, 1962)



## BRIEF REPORTS

### INTRAINDIVIDUAL VERBAL-NUMERICAL DISCREPANCIES AND PERSONALITY<sup>1</sup>

ARNOLD E. DAHLKE

*University of Minnesota*

AND

RICHARD H. DANA

*West Virginia University*

Intraindividual differences in verbal and numerical abilities have been related to personality characteristics and to psychopathology. Dana, Dahlke, and Mueller (1959) used extreme Verbal (V) and Quantitative (Q) SCAT scores; "control" subjects had smaller difference scores. Groups were compared on 22 MMPI scales; 14 of 176 *t* tests were significant and no estimation of expected number under the null hypothesis was given. Blind analysis of group profiles indicated greater subjectivity, repression, projection, and distorted thinking for extreme verbal subjects regardless of sex. These results were consistent with past research. However, percentile difference scores were used with no control for inequality of percentile units. Contrasting particular personality variables rather than entire profiles makes determination of chance expectation difficult. Percentile scores may be used with careful selection of groups. A split-plot analysis of variance design for repeated measurements employs the entire profile.

Using this design on 425 freshmen, SCAT V and Q scores were converted to percentiles, and percentile distributions were divided into three parts, V and Q low (0-29), medium (30-70), and high (71-100). Combinations resulted in six classifications with V percentile higher than Q and six with Q higher than V. Five subjects were selected randomly from each of these 12 groups. MMPI records were scored for the validity and clinical scales using *K* corrections and *T* scores. Validity and clinical scales were analyzed as

separate profiles. Profile heights were tested by the ratio of the Between Groups mean square to the Subjects Within Groups mean square; profile shapes were tested by the ratio of the Groups  $\times$  Scales mean square to the Subjects  $\times$  Scales Within Groups mean square. The Between Group mean square was broken down into a set of eight orthogonal contrasts. Each contrast was tested by the Subjects Within Groups mean square. None of the contrasts approached significance except between the extreme verbal and numerical groups for the clinical scales where the extreme verbal profile is slightly higher ( $p < .10$ ). The Groups  $\times$  Scales sum of squares was broken down into Scales  $\times$  each of the same set of contrasts and each component was tested by the Scales  $\times$  Subjects Within Groups mean square. The validity profiles of the higher verbal and numerical groups are significantly different in shape ( $p < .025$ ); the difference in shape of the clinical profiles for these groups approaches significance ( $p < .10$ ). The difference in shape between the extreme verbal and numerical groups also approaches significance ( $p < .10$ ). Validity and clinical profiles of the high numerical group in the upper percentile range and the high numerical group in the lower percentile range are significantly different ( $p < .005$ ). Blind analysis of profiles and identification of V or Q classification occurred.

Previous studies are supported. The results cannot be explained by the absolute standing of the verbal percentile since only one of three high V classifications attained significance. The sample size permits generalization to the population from which it was drawn although broader generalization is questionable.

#### REFERENCE

- DANA, R., DAHLKE, A., & MUELLER, D. Intraindividual verbal-numerical discrepancies and personality. *Amer. Psychologist*, 1959, 14, 558.

(Received May 9, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Richard H. Dana, Department of Psychology, West Virginia University, Morgantown, West Virginia, or for a fee from the American Documentation Institute. Order Document No. 7402 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

## DISCRIMINATIONS WITHIN THE DELINQUENCY CONTINUUM ON GOUGH'S SOCIALIZATION SCALE<sup>1</sup>

GAYLORD L. THORNE

*Fort Worden Diagnostic and Treatment Center, Port Townsend, Washington*

Gough (1960) claims that discriminations within the continuum of delinquency can be made using his Socialization (So) scale. The present study tested this claim in a sample of 58 delinquent boys aged 14-9 to 17-10. It was predicted that So scores would be lower on recidivists than nonrecidivists; that high So scorers commit less serious offenses than low scorers; and that high So scorers are more likely to be placed in an open institutional program where internal controls and responsibility are emphasized.

In addition, two questions of interest were examined. First, would there be any difference in So scores between "social" and "solitary" delinquents, following the research of Randolph, Richardson, and Johnson (1961). Second, does the So scale possess discriminant validity between intelligence test scores (WAIS or WISC) and between a socially desirable response set (SD).

The results showed no difference in So scores between recidivists and nonrecidivists ( $t = .92$ ), and no difference between those boys requiring a secure setting versus those placed in open settings ( $t = 1.25$ ). There was a complete absence

<sup>1</sup> An extended report of this study may be obtained without charge from G. L. Thorne, 1141 East Rose Lane, Phoenix 14, Arizona, or for a fee from the American Documentation Institute. Order Document No. 7401 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

of relationship ( $r = .001$ ) between So scores and a scale measuring severity of delinquent acts. Those getting high So scores were as likely to commit such offenses as assault, fire setting, and rape as those getting low scores.

There was a difference found between social and solitary delinquents, the latter getting lower So scores ( $t = 2.01$ ,  $p < .05$ ). This would support claims that there are reliable differences between these two types of delinquents, and it appears the dimension warrants further study and ultimately integration into a theoretical framework.

The So scores correlate only .08 with IQ scores (WAIS or WISC), but an  $r$  of .51 ( $p < .001$ ) was found with SD scores. This suggests that the concept of intelligence cannot be substituted for the present one of socialization, whereas the concept of a socially desirable response set can be substituted moderately well for that of socialization.

The present findings would not support claims that the So scale can discriminate between institutionalized delinquents. It appears that the clinician working in a delinquency institution might find it difficult to use the So scores even for screening purposes.

### REFERENCES

- GOUGH, H. G. Theory and measurement of socialization. *J. consult. Psychol.*, 1960, 24, 23-30.  
RANDOLPH, MARY H., RICHARDSON, H., & JOHNSON, R. C. A comparison of social and solitary male delinquents. *J. consult. Psychol.*, 1961, 25, 293-295.

(Received May 15, 1962)



## TYPE OF BRAIN DAMAGE AND INTELLECTUAL FUNCTIONING IN CHILDREN<sup>1</sup>

DANIEL V. CAPUTO, WILLIAM E. EDMONSTON, JR., LUCIANO L'ABATE,  
AND SAMUEL R. RONDBERG

*Washington University School of Medicine, St. Louis*

Recent studies of children with minimal brain dysfunction and with diffuse brain damage have indicated the need for a more consistent assessment of their intellectual functioning. By eliminating the influence of intellectual level from subtest patterning, through covariance analysis, one may obviate the masking of subtest differences and avoid some of the biases involved in subject selection. We attempted to test two hypotheses that bear on the relationship between type of brain damage and intellectual functioning: (a) that brain damage lowers intellectual functioning in comparison with emotional disturbance in children, children with seizures displaying less impairment than children with nonseizure brain damage, and (b) that there are differences in particular areas of intellectual functioning between two specified types of brain damage and controls.

Brain damaged subjects for this study consisted of two independent samples of children from an outpatient clinic (Sample I) and from a county health department (Sample II) who had been administered the WISC or WBI. Each experimental sample consisted of two groups of brain damaged children, a seizure group and a nonseizure group. Control subjects consisted of four groups of children referred to the same clinic and health department for emotional disturbance with no medical or historical evidence of brain damage. In terms of overall intellectual

functioning, the results tend to support our first hypothesis. Seizure groups were intermediate between control groups and nonseizure brain damage groups, although some of the differences between means are not significant. Evidence for our second hypothesis is equivocal. The findings concerning particular subtests, after covariance analysis, indicate that the seizure groups in both samples were virtually indistinguishable from their controls. Only one significant difference was found in Sample II, the seizure groups scoring significantly *higher* than controls on the Object Assembly subtest. Children with seizures then do not appear to suffer marked decrement in any specific area of intellectual functioning.

The nonseizure brain damaged groups showed a significantly lower Picture Completion score in comparison with controls in both Sample I and Sample II after covariance analysis ( $p < .05$ ). Two other subtest scores, Arithmetic and Vocabulary, were found to be significantly lower in the nonseizure brain damaged group in Sample I. The differences between our two samples may relate to at least two factors: (a) the socioeconomic level of the subjects and (b) the referring agency.

In summary, our data lend themselves to the following interpretations: (a) on Wechsler's Full Scale, Verbal and Performance IQs children grouped as "seizure brain damaged" and "nonseizure brain damaged" do not differ from one another, although both groups tend to differ from matched groups of emotionally disturbed children, (b) children with "nonseizure brain damage" yield lower Picture Completion subtest scores than matched emotionally disturbed children, and (c) Block Design and Digit Symbol subtests, while they may be quite sensitive to organic pathology, seem to be highly related to psychopathology as well.

(Received June 20, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Daniel V. Caputo, Department of Psychiatry and Neurology, Washington University School of Medicine, 4940 Audubon Avenue, St. Louis, Missouri, or for a fee from the American Documentation Institute. Order Document No. 7400 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

## JUDGMENTS OF MOTIVATION FOR PSYCHOTHERAPY: SOME FURTHER EXPLORATIONS<sup>1</sup>

MARTIN S. WALLACH

*School of Medicine, University of North Carolina*

In a previous study (Wallach & Strupp, 1960) it was observed that as patients' motivation for therapy increased, concomitant shifts were apparent in a number of dependent measures. Other investigators have also pointed to the dimension of patient motivation as being one of the most reliable ratings to correlate with a variety of other measures. This study aimed to explore further the definition and use of the concept of motivation for therapy. It is important to note that the primary concern was to elucidate the interviewers' conceptualizations of the concept.

Medical students, 71 interviewing 151 patients, and 20 psychiatric residents interviewing 182 patients completed questionnaires upon termination of the diagnostic interviews. The questionnaire called for clinical estimates, proposed treatment plans, and the interviewer's personal reaction toward the patient. Lastly, there was the estimate of the patient's motivation for psychotherapy rated on a seven-point scale and a completion type item in which the interviewer specified the cues he employed in deriving his estimate of the patient's motivation for psychotherapy.

<sup>1</sup> This research was supported by a grant (M-2171, Hans H. Strupp, Principal Investigator) from the National Institute of Mental Health, Public Health Service.

An extended report of this study may be obtained without charge from Martin S. Wallach (Department of Psychiatry, North Carolina Memorial Hospital, Chapel Hill, North Carolina) or for a fee from the American Documentation Institute. Order Document No. 7399 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$2.50 for microfilm or \$1.75 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

A factor analysis yielded two major factors, and it was demonstrated that estimates of motivation for therapy are a summary construct in the complex of impressions about a patient. However, this central role appears limited to situations in which the decision has been made to initiate therapy, and the task at hand is of evaluating the patient as a therapy participant rather than as a potential candidate.

Medical students tended to rate patient motivation as greater than did residents. This finding is believed to reflect the medical students' relative face value acceptance of patients' overt pleas for help. Psychiatric residents, on the other hand, seem to have learned that the defenses of emotionally disturbed persons tend to oppose the correction of painful symptoms or patterns of behavior, and thus they are more skeptical of overt pleas for assistance. Medical students were concerned with assessing need for help whereas residents were more set to assess the patient's ability to participate in therapy. Members of both groups equally often made reference to the same kinds of cues and dimensions in estimating motivation. The 5 (of 31 coded) categories of cues most often cited were the patient's stated desire for help, the referral source, the patient's conception of his psychiatric illness, reference to keeping the diagnostic appointment, and the patient's willingness to expend effort to get therapy.

### REFERENCE

- WALLACH, M. S., & STRUPP, H. H. Psychotherapists' clinical judgments and attitudes towards patients. *J. consult. Psychol.*, 1960, 24, 316-323.

(Received June 24, 1962)



## A FURTHER FOLLOW-UP COMPARISON OF PSYCHOTHERAPEUTIC PROGRAMS<sup>1</sup>

GEORGE W. FAIRWEATHER<sup>2</sup> AND RALPH SIMON<sup>3</sup>

*Veterans Administration Hospital, Perry Point, Maryland*

In an earlier publication (Fairweather, Simon, Gebhard, Weingarten, Holland, Sanders, Stone, & Reahl, 1960) four psychotherapy programs were compared for nonpsychotic, short-term and long-term psychotic patients. The four contrasted treatments were: a control consisting of individual work assignments; individual psychotherapy and individual work assignments; and a group living situation consisting of group activities. Comparisons were made on a number of diverse inventories, projective devices, behavior rating scales and 6-month follow-up community adjustment. This study continues the follow-up comparisons at 18 months. The questionnaire consists of nine items administered to 86 of the original 96 subjects. The items concern: (a) employment, (b) rehospitalization, (c) alcoholic indulgence, (d) antisocial behavior, (e) friendships, (f) interpersonal communication, (g) general adjustment, (h) problem behavior, and (i) degree of illness.

By the eighteenth month, no significant differences obtain among the four treatment programs in any area of community adjustment. The significantly greater percentage of patients in the three psychotherapy conditions employed during the first 6 months when compared with the work condition disappear by 18 months. On the other hand, significant differences between the

three diagnostic groups on remaining out of the hospital ( $\chi^2 = 11.45$ ,  $p < .01$ ), employment ( $\chi^2 = 9.40$ ,  $p < .01$ ), having friends ( $\chi^2 = 11.63$ ,  $p < .01$ ), and degree of illness ( $\chi^2 = 7.98$ ,  $p < .05$ ), present at 6 months continue during the next 12 months of extramural living. As in the 6-month follow-up, the most adaptive community adjustment occurs in the short-term psychotic group with long-term psychotic patients showing the poorest adjustment and the non-psychotics falling between the two. Correlations between 6- and 18-month follow-up items show that eight of the nine correlations are significant at the .01 level. They range from .93 to .53. Such high correlations suggest that the first 6 months of community adjustment are crucial and constitute a sufficient length of time to establish significant follow-up trends. In the earlier study (Fairweather, et al., 1960), only one measure correlated significantly with posttreatment community adjustment at the end of 6 months. Again, at the end of 18 months, only one measure correlates significantly with any of the follow-up criteria. The last group therapy rating correlates .50 ( $p < .05$ ) with remaining out of hospital and .49 ( $p < .05$ ), with being employed. The disappearance of significant employment differences by 18 months suggests that psychotherapy effects are of short-term duration particularly for chronic psychotics and hospitalized neurotics who return to the hospital within 18 months at the rates of 72.5 and 55.6 percent, respectively. It seems that for these groups a social system aimed at maintaining them in the community needs to be developed.

### REFERENCE

- FAIRWEATHER, G. W., SIMON, R., GEBHARD, M. E., WEINGARTEN, E., HOLLAND, J. L., SANDERS, R., STONE, G. B., & REAHL, J. E. Relative effectiveness of psychotherapeutic programs: A multi-criteria comparison of four programs for three different patient groups. *Psychol. Monogr.*, 1960, 74(5, Whole No. 492).

(Received July 6, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from George W. Fairweather, Chief, Social-Clinical Psychology Research and Service Program, Veterans Administration Hospital, Palo Alto, California, or for a fee from the American Documentation Institute. Order Document No. 7457 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> Now at Veterans Administration Hospital, Palo Alto, California.

<sup>3</sup> Now at National Institutes of Health, Bethesda, Maryland.

# ***New From Macmillan--***

## **PATTERNS OF PSYCHOPATHOLOGY: Case Studies of Behavioral Dysfunction**

*By Melvin Zax, University of Rochester and George Strickler, Goucher College*

From the files of numerous specialists and practitioners, the authors have assembled a highly readable and vitally interesting collection of case histories, over 100 in all, that demonstrate a wide variety of psychopatho-

logical phenomena. Organized into chapters representing various diagnostic syndromes, the text discusses each syndrome, follows with several pertinent cases, and concludes each chapter with integrative material.

Paper, \$3.50

## **PERSONALITY AND BEHAVIOR**

*By Jesse E. Gordon,  
University of Michigan*

Drawing on recent theoretical and research findings in both experimental psychology and clinical psychoanalysis, this well-documented volume provides a detailed analysis of personality functioning and formation.

## **PERSONAL ADJUSTMENT: An Approach Through the Study of Healthy Personality**

Second Edition

*By Sidney M. Jourard,  
University of Florida*

Here is an effective, *positive* concept of mental health. The author begins his discussion by presenting some illuminating commentary on significant modern views of personality. He then elaborates on his own theory of what constitutes a healthy personality.

## **DYNAMICS OF ADOLESCENT ADJUSTMENT**

*By Thomas F. Staton,  
Huntingdon College*

In this practical manual on adolescent behavior, Dr. Staton explains how teen-agers think and feel, why they act as they do, how to recognize when they need help, and how to help them.

## **MENTAL HYGIENE FOR TEACHERS:**

### **A Book of Readings**

*Edited by Lester D. Crow,  
Brooklyn College,  
and Alice Crow,  
formerly of Brooklyn College*

These readings provide a practical and challenging presentation of the principles of mental hygiene. The prevention of mental illness and the preservation of mental health are stressed.

## **STUDIES IN ADOLESCENCE**

*By Robert E. Grinder,  
University of Wisconsin*

Forty-five selections from the writings of authorities on psychology, sociology, and anthropology introduce a variety of viewpoints on the analysis of adolescent behavior.

## **THE PSYCHOLOGY OF ADOLESCENCE**

Second Edition

*By Arthur T. Jersild,  
Columbia University*

Dr. Jersild, continues, in this edition, his critically acclaimed analysis of the adolescent through the youngster's personal experience of his own existence.

**The Macmillan Company**

**60 Fifth Avenue**

**New York 11, N.Y.**



# Volume II

**Hans H. Strupp**  
**Lester Luborsky**  
*Editors*

of

# RESEARCH in PSYCHOTHERAPY

Proceedings of the Second Conference on Research in Psychotherapy held at Chapel Hill, N.C., in May 1961. Leading research workers in psychology and psychiatry deal with selected research issues and assess new research developments. Their efforts place research in psychotherapy more squarely within psychological research concerned with understanding interpersonal processes.

**The psychotherapist's  
contribution to the treat-  
ment process**

**Measuring personality change  
in psychotherapy**

**Definition, measurement, and  
analysis of significant  
variables**

The book is a follow up of *Research in Psychotherapy*, published by the American Psychological Association in 1959, which reported on the First Conference on Research in Psychotherapy held in Washington, D.C., in 1958.

## Contributors

Barbara J. Betz  
Jack Block  
Edward S. Bordin  
John M. Butler  
Kenneth Mark Colby  
Jerome D. Frank  
Ernest A. Haggard  
Leonard Krasner  
Henry L. Lennard  
Daniel J. Levinson  
Maurice Lorr  
Lester Luborsky  
Morris B. Parloff  
Laura N. Rice  
Irwin C. Rosen  
Nevitt Sanford  
George Saslow  
David Shakow  
John M. Shlien  
Richard S. Siegal  
William U. Snyder  
Ian Stevenson  
Hans H. Strupp  
Alice K. Wagstaff

329 pp.  
plus Indexes  
\$3.50

**Order from:**

**AMERICAN PSYCHOLOGICAL ASSOCIATION**  
1333-16th St., N.W., Washington 6, D.C.

# JOURNAL OF CONSULTING PSYCHOLOGY

June 1963

Vol. 27, No. 3

## Contents

### ARTICLES

Reliability of Retrospective Reports of Adolescence: <i>Irene Rosenthal</i> - - - - -	189
Dyadic Speech Patterns, Orientation, and Social Reinforcement: <i>Frederick H. Kanfer, Bernard M. Bass, and Irvin Guyett</i> - - - - -	199
Questionnaire Measure of the Hysteroid/Obsessoid Component of Personality; The HOQ: <i>T. M. Caine and L. G. Hawkins</i> - - - - -	206
Deviant Response Patterns as a Function of Chronological Age: <i>S. O. Hesterly</i> - - - - -	210
Construction and Validation of the Food Attitude Scale: <i>Donn Byrne, Carole Golightly, and E. J. Capaldi</i> -	215
Verbal Conditioning of Schizophrenics and Normals in a Situation Resembling Psychotherapy: <i>Joan Slechta, William Gwynn, and Crocker Peoples</i> - - - - -	223
Factor Scales for the California Psychological Inventory: <i>Robert C. Nichols and Richard R. Schnell</i> - -	228
Sibling Patterns and Social Adjustment among Normal and Psychiatrically Disturbed Soldiers: <i>William Vogel and Carl G. Lauterbach</i> - - - - -	236
Scores of Psychotic Patients on the Maudsley Personality Inventory: <i>T. G. Crookes and S. J. Hutt</i> - -	243
Effects of Anger Arousal and Similarity upon the Attribution of Hostility to Pictorial Stimuli: <i>Seymour Feshbach, Robert D. Singer, and Norma Feshbach</i> - - - - -	248
Simulation on the California Psychological Inventory and the Adjustment of the Simulator: <i>Francis M. Canter</i> - - - - -	253
Effect of Professional Training and Amount and Accuracy of Information on Behavioral Prediction: <i>Janis H. Weiss</i> - - - - -	257

### NOTES AND COMMENTS

Applicability of the Norms of the Wechsler Intelligence Scale for Children to Five-Year-Olds: <i>Michael Beldoch</i> - - - - -	263
Research Creativity in Psychology Graduate Students: <i>Martha T. Mednick</i> - - - - -	265
MMPI Profiles of Epileptics: A Further Evaluation: <i>Edward J. Jordan, Jr.</i> - - - - -	267
Sex and Personality Differences in Relation to Fantasy: <i>C. W. LaGrone</i> - - - - -	270
Relationship between Time Imagery and Rorschach Human Movement Responses: <i>Ronald B. Kurz</i> - -	273

### BRIEF REPORTS

Characteristics of Homosexually Involved Incarcerated Females: <i>William G. Miller and Thomas E. Hannum</i> - - - - -	277
Fantasized Danger as a Function of Parent-Child Controlling Practices: <i>Donald W. Tiffany and Franklin C. Shontz</i> - - - - -	278
Cross-Validation of Stein's Symbol-Gestalt Test for Brain Damage: <i>Oscar A. Parsons, Harriet I. Maslow, and Kenneth D. Stewart</i> - - - - -	279
Effect of Intellectual Level and Neuropsychiatric Status on the Diversity of Intensity of Semantic Differential Ratings: <i>Charles Neuringer</i> - - - - -	280
Digit Span Performance as a Function of Noxious Stimulation: <i>Sandra Pyke and Neil McK. Agnew</i> - -	281
Further Support of Some Findings about the Characteristics of Smokers and Nonsmokers: <i>Bruce C. Straits and Lee Sechrest</i> - - - - -	282

## Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

### Specifications

**Brief Report.** The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the

author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

<sup>1</sup> An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Documentation No. — from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$— for microfilm or \$— for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

**Extended report.** Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (a) The extended report should be typed with single spacing for economy in duplication. (b) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (c) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.



## RELIABILITY OF RETROSPECTIVE REPORTS OF ADOLESCENCE

IRENE ROSENTHAL

*Institute of Human Development, University of California, Berkeley*<sup>1</sup>

Accuracy of memories regarding relation with parents, and social and emotional characteristics of adolescence, were studied for 100 Ss of a longitudinal study. Interviews, tests, and ratings had been obtained for the Ss as adolescents and as adults of age 38-40. Product-moment correlations were computed between retrospective, adolescent, and adult variables. Partial correlations were computed between retrospective and adolescent variables, holding adult measures constant, when memories correlated significantly with adolescent data, and also with adult measures. Men's retrospective account of adolescence represented adolescent self-pictures and observers' ratings during adolescence with a considerable degree of accuracy, while the women's memories reflected adolescence to a lesser extent. In some cases, the women's memories were more highly related to the present than the past, and in others they were associated with adolescent items other than the corresponding measures.

Reports of memories of childhood experiences have served as a basis for the development of theories of personality formation, and of techniques of psychotherapy. Verification of the accuracy of these memories, however, has been limited to the studies by Pyles, Stolz, and Macfarlane (1935) and Haggard, Brekstad, and Skard (1960), both of which investigated mother's anamnestic reports concerning the birth and early development of their children. Empirical studies of the accuracy of memories regarding the types of material obtained in psychotherapeutic interviews, such as relationship with parents, and social and emotional characteristics of childhood and adolescence, are nonexistent at the present time.

Although theories have dealt primarily with relations between early childhood and subsequent personality development, some theoretical formulations have involved the adolescent period. An opportunity to investigate the accuracy of adults' memories of adolescence was presented in the data gathered in connection with the Oakland Growth Study. The subjects of this longitudinal study, who had been studied extensively during adolescence, participated in a follow-up study which gathered material on their present lives, and their memories of adolescence.<sup>2</sup>

Thus, retrospective accounts of adolescence, obtained after a lapse of 20 years, could be compared with data which were obtained during the adolescent years.

### METHOD

#### *Subjects*

The Oakland Growth Study began in January 1932, with subjects who were then in the fifth grade of elementary school. These subjects were studied intensively during the next 7 years until their graduation from high school. In the years 1958-60, when their average age was 38-40, many of the subjects of the original study participated in testing sessions, and in a series of interviews averaging 8-12 hours in length for each subject. Detailed descriptions of the study have been published by Jones (1938, 1939, 1959).

The present study involved 48 men and 52 women who were members of the Oakland Growth Study during their last 2 years of high school, and who participated in the follow-up study. Since all subjects did not participate in both the interviews and testing sessions, the number of subjects studied in connection with any one variable was less than the total.

#### *Adolescent Variables*

The adolescent variables of the present study consisted of observer's ratings of the subjects, made during their last 2 years of high school, and the

<sup>2</sup> Data for the present study, which is part of a long-term longitudinal project of the Institute of Human Development, University of California, were gathered by Harold E. Jones under a grant from the Ford Foundation.

<sup>1</sup> Now at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

TABLE 1  
CONSISTENCY OF QUESTIONNAIRES OVER 1 YEAR  
AT THE END OF HIGH SCHOOL

Questionnaire	Product-moment correlation	
	Men	Women
5. Has fun	.43	.55
6. Is a leader	.58	.74
7. Likes to be alone, imagines	.75	.98
8. Parents like the subject's friends	.73	.35
9. Parents treat the subject as child	.39	.86
10. Not punished by parents	.70	.79
11. Doesn't mind parents	.58	.82
12. Wishes-Parents	.60	.43
13. Wishes-Neutral	.44	.33
14. Fears	.63	.66
15. Tensions	.67	.71

Note.—All correlations are significant at the two-tailed .05 level.

subjects' responses to questionnaires administered during the same period.

**Ratings.** The adolescent ratings, based on observations in a free social situation, were made independently on a 5-point scale by three observers. For the present study, the average of the two sets of ratings, made during the last 2 years of high school, was used. These ratings are described fully in an article by Newman (1946) where the reliabilities, ranging from .56-.78 in terms of product-moment correlations, are given. The ratings with assigned numbers, were: Leader (1), Uninhibited (2), Relaxed (3), and Cheerful (4).

**Questionnaires.** The study utilized questionnaires administered during the last 2 years of high school. Since reliabilities were not available for the questionnaires, product-moment correlations between the measures, given a year apart in adolescence, are given in Table 1 to provide an indication of the reliability.

1. Seven items, drawn from the Tryon (1933) Adjustment Inventory, gave descriptions of behavior, and asked the subject to rate himself on a 5-point scale ranging from "yes" to "no," as to how much he resembled the boy (or girl) described in the statement. (For example, the first item, Variable 5, consisted of the Statement "A. is a boy who seems to have a lot of fun.") An average was computed for the two adolescent responses to each of these items for the last 2 years of high school. Abbreviations of the questionnaire items, and the numbers assigned to the variables were: Subject has fun (5), Subject is a leader (6), Subject likes to be alone, imagines (7), Parents like subject's friends (8), Parents treat subject as a child (9), Subject is almost never punished (10), and Subject doesn't want to mind parents (11).

2. Two adolescent variables, "Wishes-Parents" (12) and "Wishes-Neutral" (13) were derived from a test where the subjects were asked to check things they "wished for." The score, computed for each of the last 2 years of high school, was the total number of items of each type checked.

The items, comprising the Wishes-Parents<sup>a</sup> (12) test, dealt with the subject's wish to have fewer conflicts with his parents, to receive more attention from his parents, and to have more cheerful parents.

The Wishes-Neutral (13) test consisted of five items not dealing with parents. Included in this test were wishes concerning recreation, playmates, or school. For example, one of the items of the Wishes-Neutral test was "I wish I could go swimming every Saturday." The Wishes-Neutral score was used in the study to determine whether the Wishes-Parents score represented a general tendency for the subject to change his life rather than a specific desire for changes in the relation with his parents.

3. Subjects were asked to check things they feared from a list of 61 items which concerned fears of things such as injuries to self, high places, specific persons, being lost, or being accosted (Tryon, 1939, lists these fears). The score for the test, which was the sum of the items checked, was computed for each of the last 2 years of high school. The designation for this variable with its assigned number was: Fears (14).

4. The "Tensions" test, which is a part of the Tryon (1933) Adjustment Inventory, listed anxieties, moods, and tensions, and asked the subject to reply yes or no in terms of the applicability of the item to himself. The score was obtained by finding the sum of the yes responses. In this test, the subject was asked whether he was unhappy, lonely, nervous, easily upset, and other questions of a similar nature. Scores were obtained for each of the last 2 years of high school. The name used for the test with the assigned number was: Tensions (15).

5. A score for social adjustment was obtained by summing responses (made on a 5-point scale) to those items of a longer questionnaire which dealt with social relations, and which intercorrelated significantly with each other in adolescence. The items dealt with the subject's feelings of bashfulness, difficulties in making friends, and hesitancy in talking to strangers.

<sup>a</sup> To obtain data on the concurrent validity of the retrospective Wish-Parents test, product-moment correlations were computed between the score on this test and a rating of "counter identification with parent" which denoted that the subject had stated that he had chosen a way of life opposite from the parent of the same sex. The correlations were .60 for the men, and .52 for the women. This rating, devised by Suzanne Reichard and based on the follow-up interviews, is part of a study currently being done by John S. Hatfield and Albert M. Shapiro.



<sup>5</sup> A comprehensive analysis of "Personality Continuity in the Oakland Growth Study" is in progress under the direction of John A. Clausen. A study of Tuddenham (1959) of constancy of personality ratings, utilizing part of the present sample, is not comparable to the present study both because the adolescent data were drawn from the time that the subjects were in junior high school, and because the adult ratings were based on shorter and less comprehensive interviews.



TABLE 2  
VARIABLES IN THE CORRELATION MATRIX

Adolescent	Adult
<p>Ratings</p> <ol style="list-style-type: none"> <li>1. Leader</li> <li>2. Uninhibited</li> <li>3. Relaxed</li> <li>4. Cheerful</li> </ol> <p>Questionnaires</p> <ol style="list-style-type: none"> <li>5. Subject has fun</li> <li>6. Subject is a leader</li> <li>7. Subject likes to be alone, imagines</li> <li>8. Parents like subject's friends</li> <li>9. Parents treat subject as child</li> <li>10. Subject is not punished by parents</li> <li>11. Subject doesn't mind parents</li> <li>12. Wishes-Parents</li> <li>13. Wishes-Neutral</li> <li>14. Fears</li> <li>15. Tensions</li> <li>16. Social adjustment</li> </ol>	<p>Ratings</p> <ol style="list-style-type: none"> <li>17. Leader</li> <li>18. Uninhibited</li> <li>19. Relaxed</li> <li>20. Cheerful</li> </ol> <p>Retrospective ratings</p> <ol style="list-style-type: none"> <li>21. Leader</li> <li>22. Uninhibited</li> <li>23. Relaxed</li> <li>24. Cheerful</li> </ol> <p>Questionnaires answered in terms of the present</p> <ol style="list-style-type: none"> <li>25. Fears</li> <li>26. Tensions</li> <li>27. Social adjustment</li> </ol> <p>Retrospective questionnaires</p> <ol style="list-style-type: none"> <li>28. Subject has fun</li> <li>29. Subject is a leader</li> <li>30. Subject likes to be alone, imagines</li> <li>31. Parents like subject's friends</li> <li>32. Parents treat subject as child</li> <li>33. Subject is not punished by parents</li> <li>34. Subject doesn't mind parents</li> <li>35. Wishes-Parents</li> <li>36. Wishes-Neutral</li> <li>37. Fears</li> </ol>

each sex while correlations involving questionnaires had sample sizes ranging from 34-43 for the men, and 31-39 for the women.

All adolescent variables covered the last 2 years of high school except the test, "Social adjustment," for which only 1 year's results were available. For the adolescent questionnaire items, and ratings (which were scores on a single variable), scores were composited for the last 2 years of high school before the correlations were calculated. For the tests whose scores were sums of a number of items (i.e., the Tensions, Fears, and Wishes tests), the scores for each of the last 2 years of high school were included in the matrix of data used to obtain correlations, and the two correlations for each variable were then averaged to obtain the value for the correlation used in the study.

*Relations between Memories of Adolescence and Adolescent and Adult Measures of Personality*

*Memory of Relation to Parents during Adolescence*

As shown in Table 3, responses on the retrospective questionnaires dealing with parental relation correlated significantly with responses on several adolescent measures of parental relation in the case of the men, but to only a few for the women. Each of the men's retrospective questionnaires, except the one dealing with punishment, was correlated significantly with the identical questionnaire of adolescence. For the women, the two items dealing with memory of discipline were unrelated to any adolescent measure, and the remaining questionnaires each correlated with one questionnaire dealing with adolescent parental attitudes, but not to the questionnaire which was identical.

To determine whether reports of memories were influenced by the present rather than the past, correlations had been computed between retrospective and adult measures (Table 3). The men's reports of memories of relation to parents were related significantly to some adult variables suggesting the possibility that aspects of their adult personalities might be distorting their memories of their parents. To investigate this possibility, partial correlations were computed between the retrospective parental variables and the adolescent parental variables with the adult measures held constant. These partial correlations deviated by less than .05 from the original correlations except for the correlation between "Parents treat subject as child" and "Not punished" which became .19 with adult Tensions held con-

stant. Thus, the correlations obtained for the men between parental retrospective variables and adolescent parental variables could not be explained on the basis of adult personality measures.

Inspection of Table 3 reveals that the men's retrospective questionnaires were associated with adolescent variables other than those dealing with parental relations. This suggests the hypothesis that the men's memories of adolescent parental relation might reflect a general picture of their adolescent life rather than the specific parental relation. Another explanation of the significant correlations obtained between retrospective parental variables and adolescent non-parental-variables might be, however, that both the memories and the adolescent variables were related to some adult variables.

TABLE 3  
RELATION BETWEEN MEASURES OF MEMORIES OF ADOLESCENT PARENTAL RELATIONSHIP  
AND MEASURES OF ADOLESCENT AND ADULT PERSONALITY

Retrospective questionnaire	Adolescent variables	Product-moment $r$	Adult variables	Product-moment $r$
Men				
31. P like friends—Q	8. P like friends—Q	.54	26. Tensions—Q	-.39
	5. Has fun—Q	.39	25. Fears—Q	-.36
	9. P treat as child—Q	-.35		
32. P treat as child—Q	12. Wishes-Parents—Q	.56	26. Tensions—Q	.48
	15. Tensions—Q	.46		
	9. P treat as child—Q	.36		
	10. Not punished—Q	-.33		
33. Not punished—Q	5. Has fun—Q	.32	20. Cheerful—R	.39
			17. Leader—R	.37
34. Doesn't mind P—Q	11. Doesn't mind P—Q	.52		
	6. Leader—Q	.40		
35. Wishes-Parents—Q	9. P treat as child—Q	.59		
	12. Wishes-Parents—Q	.46		
Women				
31. P like friends—Q	12. Wishes-Parents—Q	-.41		
32. P treat as child—Q	8. P like friends—Q	-.39		
33. Not punished—Q			20. Cheerful—R	-.47
34. Doesn't mind P—Q			18. Uninhibited—R	.45
35. Wishes-Parents—Q	8. P like friends—Q	-.45		

Note.—P denotes parents; R denotes rating; Q denotes questionnaire.  
All correlations are significant,  $p < .05$ , two-tailed.

To determine whether the correlations between the retrospective parental questionnaires and the adolescent nonparental variables could be explained on the basis of adult personality measures, partial correlations were computed between the retrospective and the adolescent variables with the adult variables held constant. These partial correlations were not significant. (For Parents Like Subject's Friends versus Subject has Fun, and Parents Treat Subject as Child versus Tensions, the correlations became .26 and .28, respectively, with adult Tensions held constant. The correlation between Not Punished and Subject had Fun became .18 with Cheerful held constant.) The fact that the correlations of the retrospective parental variables with adolescent nonparental variables could be attributed to the effect of adult variables while those with parental adolescent variables could not, indicates that memories of adolescent parental relation probably reflected the actual adolescent parental relation. Further evidence that memories of adolescent parental variables were related specifically to parental adolescent relation came from the results obtained with the Wishes-Neutral test. The score on the retrospective form of this test correlated with the corresponding score on the adolescent test ( $r = .44$ ) but not with the adolescent Wishes-Parents test, just as the retrospective form of the latter test correlated only with the corresponding adolescent measure.

For the women, there was less evidence that their responses were determined by adolescent relation with their parents. Although the retrospective items tended to be related to adolescent parental items rather than to nonparental variables, they did not correlate significantly with the identical items, and the number of correlates was low. In addition, one of the retrospective questionnaires, Wishes-Parents, was related to the adult variable "Uninhibited" as well as to the adolescent variable. To determine whether the adult variable Uninhibited could account for the correlation between the retrospective Wishes-Parents and the adolescent Wishes-Parents tests, the partial correlation was computed with adult Uninhibited

held constant. This partial correlation was .30, a nonsignificant value.<sup>6</sup>

*Adolescent Memories—Dealing with  
Aspects of Adolescence other than  
Parental Relationship*

Tables 4 and 5 present the adolescent and adult variables which correlated significantly with retrospective measures of social and emotional characteristics of adolescence. (The retrospective questionnaire, Fears, which is not included in the tables, will be discussed separately in a subsequent section.) For both sexes, memories were fairly highly related to data obtained during adolescence.

The retrospective questionnaires, Has Fun, Leader, and Alone, Imagines, had significant correlations with the identical adolescent variables (Table 5). Two of the retrospective ratings, Leader and Uninhibited, correlated significantly with the corresponding adolescent variable for both sexes (Table 4). The retrospective rating for the trait, Cheerful, correlated significantly with adolescent Cheerful for the men, but not for the women. The retrospective rating, Relaxed, failed to correlate with the corresponding adolescent trait for both men and women. The fact that some of the correlations between the retrospective ratings and the adolescent ratings were insignificant might be explained by the different nature of these two sets of ratings. Thus, the retrospective ratings were self-pictures of the subjects, while the adolescent ratings were observers' impressions.

Significant correlations obtained between retrospective variables and adolescent variables could not be attributed to accuracy of memory without eliminating the effect of adult personality on the reports of adolescent memories. The fact that significant correlations had been obtained between memories and adult measures of personality, suggested the possibility that memories were being distorted to be compatible with the

<sup>6</sup> The women's Wish-Neutral test behaved in a manner similar to the Wish-Parent test in failing to correlate with the identical adolescent test. The only significant correlation of this test with an adolescent variable was with the item "Doesn't mind parents" ( $r = .51$ ).



TABLE 4

RELATION BETWEEN RATINGS OF MEMORIES OF ADOLESCENCE OTHER THAN THOSE DEALING  
WITH PARENTS, AND MEASURES OF ADOLESCENT AND ADULT PERSONALITY

Retrospective rating	Adolescent variables	Product- moment <i>r</i>	Adult variables	Product- moment <i>r</i>
Men				
21. Leader—R	1. Leader—R	.38	18. Uninhibited—R	.31
	2. Uninhibited—R	.38		
	6. Leader—Q	.36		
	4. Cheerful—R	.33		
22. Uninhibited—R	2. Uninhibited—R	.53	18. Uninhibited—R	.49
	16. Social adjustment—Q	.45	27. Social adjustment—Q	.47
	1. Leader—R	.43		
	5. Fun—Q	.39	26. Tensions—Q	-.44
	4. Cheerful—R	.39	17. Leader—R	.30
23. Relaxed—R	5. Fun—Q	.42	26. Tensions—Q	-.50
24. Cheerful—R	16. Social adjustment—Q	.52	20. Cheerful—R	.38
	5. Fun—Q	.43	18. Uninhibited—R	.32
	4. Cheerful—R	.40		
	2. Uninhibited—R	.36		
	3. Relaxed—R	.34		
	1. Leader—R	.33		
Women				
21. Leader—R	2. Uninhibited—R	.46	17. Leader—R	.67
	1. Leader—R	.44	18. Uninhibited—R	.59
	6. Leader—Q	.41	27. Social adjustment—Q	.57
			26. Tensions—Q	.43
22. Uninhibited—R		.59	18. Uninhibited—R	.71
	1. Leader—R	.55	17. Leader—R	.55
	2. Uninhibited—R	.49	20. Cheerful—R	.37
	5. Fun—Q	.42	27. Social adjustment—Q	.34
	6. Leader—Q	.39		
	4. Cheerful—R	.37		
	8. P like friends—Q	.35	20. Cheerful—R	.30
23. Relaxed—R	1. Leader—R	.37	20. Cheerful—R	.54
24. Cheerful—R	1. Leader—R		18. Uninhibited—R	.47
			27. Social adjustment—Q	.43
			17. Leader—R	.33

Note.—P denotes parents; R denotes rating; Q denotes questionnaire. All correlations are significant,  $p < .05$ , two-tailed.

adult life situation. To investigate this possibility, partial correlations were computed between retrospective and adolescent variables holding adult measures constant. Since the results (Table 6) followed a different pattern for the men than for the women, they will be discussed separately.

For the men, all partial correlations (Table 6) were significant except for the partial correlation between the retrospective rating

Relaxed and the adolescent questionnaire Has Fun with adult Tensions held constant. Thus, the relation found for the men between retrospective and adolescent variables could be attributed to accuracy of memory.

The partial correlations obtained for the women present an inconsistent picture of the effect of adult personality on memories of adolescence. The retrospective rating and the retrospective questionnaire for the trait,

TABLE 5

RELATION BETWEEN QUESTIONNAIRE MEASURES OF MEMORIES OF ADOLESCENCE, OTHER THAN THOSE DEALING WITH PARENTS, AND MEASURES OF ADOLESCENT AND ADULT PERSONALITY

Retrospective questionnaire	Adolescent variables	Product-moment $r$	Adult variables	Product-moment $r$
Men				
28. Fun—Q	16. Social adjustment—Q	.42		
	1. Leader—R	.38		
	4. Cheerful—R	.37		
	5. Fun—Q	.37		
	2. Uninhibited—R	.35		
	9. P treated as child—Q	-.32		
29. Leader—Q	6. Leader—Q	.41		
30. Alone, Imagines—Q	7. Alone, Imagines—Q	.55	27. Social adjustment—Q	-.38
	9. P treat as child—Q	.46	18. Uninhibited—R	-.34
	4. Cheerful—R	-.42	26. Tensions—Q	.32
	2. Uninhibited—R	-.41		
	1. Leader—R	-.41		
	15. Tensions—Q	.34		
Women				
28. Fun—Q	1. Leader—R	.47	27. Social adjustment—Q	.41
	5. Fun—Q	.37		
	2. Uninhibited—R	.36		
29. Leader—Q	6. Leader—Q	.64	17. Leader—R	.36
	1. Leader—R	.44	18. Uninhibited—R	.35
	2. Uninhibited—R	.42	27. Social adjustment—Q	.35
30. Alone, Imagines—Q	10. Not punished—Q	.50		
	9. P treat as child—Q	.37		
	7. Alone, Imagines—Q	.36		

Note.—P denotes parents; R denotes rating; Q denotes questionnaire.  
All correlations are significant,  $p < .05$ , two-tailed.

Leader, and the retrospective rating for Uninhibited had insignificant partial correlations with adolescent ratings but had significant partial correlations with adolescent questionnaires. For the retrospective questionnaire, Has Fun, and the retrospective rating, Cheerful, all correlations became insignificant when adult measures were held constant. For the retrospective variable Relaxed, however, the partial correlation with the adolescent rating Leader, holding the adult rating Cheerful constant, was significant. Since no consistent picture emerged from these partial correlations, the question whether adult personality affects memories of social and emotional characteristics for

women remains unanswered in the present study.

The last retrospective questionnaire to be discussed, the Fears test, differed from the questionnaires presented in the preceding sections both in content and in relations with adolescent and adult measures. The retrospective form of this test was not correlated significantly with the adolescent test, but was highly related to the adult test for both men ( $r = .64$ ) and women ( $r = .75$ ). The high correlation between the adult and retrospective forms of the Fears might be due partially to the fact that the two forms were given successively in the testing session.

## DISCUSSION

The men's memories of their relationship with their parents, and of their social and emotional characteristics during adolescence were found to be significantly related to adolescent self-pictures, and to observers' ratings obtained during adolescence. For the women, however, the results of the study did not give a consistent picture of the relation between memory and actual data of adolescence.

The lack of association found for the women for the retrospective measures of parental relations, as contrasted with the considerable association found between memory and adolescence for nonparental variables, indicated that inaccuracy in memory was related specifically to the area of parental relation. Since the number of parental measures was low, however, this can only be a tentative conclusion. In addition, the fact that the relation between the

TABLE 6  
PARTIAL CORRELATIONS BETWEEN RETROSPECTIVE AND ADOLESCENT  
VARIABLES WITH ADULT VARIABLES HELD CONSTANT

Variables being correlated <sup>a</sup>		Adult variable held constant <sup>a</sup>	Partial correlation
Retrospective	Adolescent		
Men			
30. Alone, Imagines—Q	7. Alone, Imagines—Q	27. Social adjustment—Q	.53*
30. Alone, Imagines—Q	7. Alone, Imagines—Q	18. Uninhibited—R	.51*
30. Alone, Imagines—Q	7. Alone, Imagines—Q	26. Tensions—Q	.52*
30. Alone, Imagines—Q	4. Cheerful—R	27. Social adjustment—Q	.36*
30. Alone, Imagines—Q	4. Cheerful—R	18. Uninhibited—R	.34*
30. Alone, Imagines—Q	4. Cheerful—R	26. Tensions—Q	.35*
30. Alone, Imagines—Q	1. Leader—R	18. Uninhibited—R	.32*
21. Leader—R	6. Leader—Q	18. Uninhibited—R	.32*
21. Leader—R	2. Uninhibited—R	18. Uninhibited—R	.40*
22. Uninhibited—R	2. Uninhibited—R	27. Social adjustment—Q	.39*
22. Uninhibited—R	2. Uninhibited—R	26. Tensions—Q	.53*
22. Uninhibited—R	2. Uninhibited—R	18. Uninhibited—R	.46*
22. Uninhibited—R	16. Social adjustment—Q	27. Social adjustment—Q	.50*
22. Uninhibited—R	16. Social adjustment—Q	26. Tensions—Q	.42*
22. Uninhibited—R	16. Social adjustment—Q	26. Tensions—Q	.26
23. Relaxed—R	5. Has fun—Q	20. Cheerful—R	.47*
24. Cheerful—R	16. Social adjustment—Q	20. Cheerful—R	.30*
24. Cheerful—R	4. Cheerful—R		
Women			
28. Has fun—Q	1. Leader—R	27. Social adjustment—Q	.32
28. Has fun—Q	5. Has fun—Q	27. Social adjustment—Q	.28
29. Leader—Q	6. Leader—Q	17. Leader—R	.59*
29. Leader—Q	6. Leader—Q	27. Social adjustment—Q	.57*
29. Leader—Q	1. Leader—R	27. Social adjustment—Q	.32
21. Leader—R	1. Leader—R	17. Leader—R	.26
21. Leader—R	6. Leader—Q	17. Leader—R	.37*
21. Leader—R	6. Leader—Q	27. Social adjustment—Q	.36*
21. Leader—R	6. Leader—Q	26. Tensions—Q	.40*
21. Leader—R	6. Leader—Q	18. Uninhibited—R	.25
22. Uninhibited—R	2. Uninhibited—R	18. Uninhibited—R	.42*
22. Uninhibited—R	5. Has fun—Q	27. Social adjustment—Q	.43*
22. Uninhibited—R	5. Has fun—Q	20. Cheerful—R	.32*
23. Relaxed—R	1. Leader—R	17. Leader—R	.28
24. Cheerful—R	1. Leader—R		

<sup>a</sup> R denotes rating; Q denotes questionnaire.

<sup>b</sup>  $p < .05$ , two-tailed.



women's retrospective and adolescent ratings could be attributed to a considerable extent to adult personality leaves fewer data than were available for the men on which to base the conclusion that nonparental memories tended to be accurate. (Although the relations between retrospective and adolescent ratings could be attributed to adult personality, the interpretation that memories were accurate was not excluded by the results.)

### *Comparison with Previous Studies*

Adults' memories of adolescence in the present study corresponded more closely to the past than mothers' memories concerning the birth and early development of their children, as studied by Pyles et al. (1935) and Haggard et al. (1960). The results of the previous studies suggested that retrospective reports would have been less accurate in the present study than in the other studies since, in the previous studies, memories for facts were more accurate than for measures such as attitudes, wishes, and anxieties, which are more comparable to the measures of the present study. There are, however, important differences between the study reported here and the other studies. First, the memories in this study concerned the subject himself, rather than his children. Secondly, both previous studies dealt only with women, and the results of the present study indicate that women's memories are related less to the past than the memories of the men.

Unlike the previous studies of anamnestic reports, the present demonstration of some validity of memories lends some support to

theories of adolescent personality development which have been based to a large extent on retrospective interviews and questionnaires. The demonstration, however, of inaccuracy in memories in some areas, such as that of the adolescent parent relation in the case of the women, suggests caution in using retrospective data both in the formulation of theories of personality, and in studies which attempt to relate adult personality characteristics to adolescent development.

### REFERENCES

- HAGGARD, E. A., BREKSTAD, A., & SKARD, A. G. On the reliability of the anamnestic interview. *J. abnorm. soc. Psychol.*, 1960, 61, 311-318.
- JONES, H. E. The California Adolescent Growth Study. *J. educ. Res.*, 1938, 31, 561-567.
- JONES, H. E. Procedures of the Adolescent Growth Study. *J. consult. Psychol.*, 1939, 3, 177-180.
- JONES, H. E., MACFARLANE, JEAN W., & EICHORN, DOROTHY H. A progress report on growth studies at the University of California. *Vita Humana*, 1959, 3, 17-31.
- NEWMAN, F. B. The adolescent in social groups: Studies in the observation of personality. *Appl. Psychol. Monogr.*, 1946, 9, 1-94.
- PYLES, MARJORIE K., STOLZ, H. R., & MACFARLANE, JEAN W. The accuracy of mothers' reports on birth and development data. *Child Developm.*, 1935, 6, 165-176.
- TRYON, CAROLINE M. A study of social and emotional adjustments: The internal consistency and inter-relationships of measurements by different techniques. Unpublished doctoral dissertation, University of California, 1933.
- TRYON, C. M. *U. C. Inventory I: Social and emotional adjustment*. Berkeley: Univer. California, Institute of Child Welfare, 1939.
- TUDDENHAM, R. D. The constancy of personality ratings over two decades. *Genet. Psychol. Monogr.*, 1959, 60, 3-29.

(Received September 1, 1962)

## DYADIC SPEECH PATTERNS, ORIENTATION, AND SOCIAL REINFORCEMENT<sup>1</sup>

FREDERICK H. KANFER

*University of Oregon Medical School*

BERNARD M. BASS

*University of Pittsburgh*

AND IRVIN GUYETT

*Louisiana State University*

48 female Ss ranked photos before and after discussion with a male partner (P) whose comments were standardized. S controlled the verbal exchange. Greater opinion changes were found when P disagreed than when he agreed. Social orientation did not affect the degree of this change. All Ss allotted more time to P than to themselves. Self-oriented Ss talked most, task-oriented Ss talked less, and interaction-oriented Ss talked least. P's agreement or disagreement did not affect talk-time differentially. Disagreement did not change Ss' speech patterns over trials. Agreement increased talk in task and interaction-oriented Ss over trials while self-oriented Ss talked less. Postexperimental questions yielded differences in attitudes among groups toward the experiment and P.

Dyadic research is of special interest to clinical psychologists for its potential contributions to a better understanding of the clinical process. The present experiment endeavors to link verbal conditioning research with studies on the effects of reinforcement on participation in small groups in order to study the role of a subject as both speaker and listener in a controlled dyadic interaction.

In most studies on the effects of reinforcement on verbal behavior, the subject's range of effective verbal behaviors is limited and closely controlled by the experimenter. The usual dyadic situation differs in that it permits a speaker to affect the behavior of the reinforcing partner in at least two ways. The speaker may obtain increased numbers of reinforcements when he increases the proportion of a critical response class, changing the content of his speech but not the overall volume of his verbal output (Kanfer & McBrearty, 1962). A second strategy, often ob-

served in two-person interactions, consists of solicitation of reinforcing statements from the listener. These verbal behaviors, classified as demands by Skinner (1957), may be manifested by the speaker's seeking of reassurance, or confirmation of his opinions. Such "reinforcement-soliciting" behaviors may be subject to modification by the same variables as speech content (opinions), or they may relate to relatively invariant personality characteristics.

From a viewpoint of group dynamic theory, the progression toward solution of a group problem also takes on importance. When the subject is exposed to comments in which the partner states agreement about the subject's opinions, the subject's attempts to influence are reinforced. Especially if both participants are of equal peer ratings and a specific task is to be solved, a rational model of group interaction suggests that the subject is encouraged to communicate more frequently and for longer durations if his partner approves his efforts (Bass, 1960, Ch. 7). While reinforcement theories stress the social and emotional effects, group dynamic theory emphasizes the value of contributions toward a solution.

The purpose of this study was to investigate the effects of social reinforcement, and

<sup>1</sup> This study was supported in part by Research Grant M-3700 (C1) from the National Institute of Mental Health, U. S. Public Health Service, F. H. Kanfer, principal investigator. The data were collected at Louisiana State University in Spring 1961. The authors wish to thank O. Hubert Campbell and Donald S. Leventhal for their assistance in the analysis of the data.

the subject's social orientation on (a) dyadic speech patterns, and (b) verbal content (opinions) when the partner's opinions were predetermined, and the subject was given full control over frequency and durations of the partner's contributions. The relationships under investigation were studied by examination of the changes in the subject's opinion after a series of interactions and the amount of time which the subject allotted to himself and his partner during the verbal exchange, for the subjects whose partners agreed or disagreed with their statements, and for the subjects who differed on Orientation Inventory scores (ORI) (Bass, 1962).

### METHOD

#### Subjects

The subjects in this study were 48 volunteer female undergraduates.

#### Apparatus

The experimental room is shown in Figure 1. On the table before the subject were two keys. The left key was marked *Talk* and the right key was marked *Listen*. Plywood partitions separated the subject from her partner, who was the experimenter's confederate. On the partner's side a slide projector was on the table. When a picture was shown, the projection surface was a 25 × 25 inch area on a light green wall, approximately 6 feet from the subject. Below the projection area two lights were mounted. A red light was wired to the key marked *Talk* and a green light was connected to the subject's *Listen* key. The subject had a free view of the picture but she could not see the partner or the experimenter. Behind the experimenter's partition a large clock, accurate to .5 second, timed each trial. Two smaller clocks were connected in series with the subject's *Talk* and *Listen* keys and their respective lights. Adjacent to the smaller clocks were two signal lights, in series with the keys. The small clocks cumulated the time on each key for each trial. In this way the experimenter recorded duration of each key depression, frequency of key depressions, and duration of silence (when neither key was depressed).

#### The Experimental Pictures

Five experimental pictures of faces were chosen from those used by Stritch (1954). A trial consisted of a 4-minute period for discussion of the face. At the beginning of each trial the subject ranked 10 adjectives typed with regard to the face. At the end of each trial the subject ranked the same adjectives again. The rank difference correlation obtained for each subject served as an indication of the subject's shift as a result of discussion with the partner. The adjectives had been used in previous research by Campbell (1961) and were

taken from Osgood's list of traits. They were: intelligent, friendly, relaxed, mature, trusting, adaptable, thrifty, fair, sociable, and calm. The sequence of picture presentations over five trials was randomized for each subject.

#### Orientation of the Subjects

Bass (1960, Ch. 8) proposed three ideal types of group members: self-oriented, interaction-oriented, and task-oriented. He argued that these types would be reinforced by different effects. The task-oriented subject would behave in a manner directed toward getting the job done; attempts to influence others toward group effectiveness would be reinforced. The interaction-oriented subject would direct his attention to maintaining conflict-free harmonious interactions; behavior producing pleasant interaction would be reinforced. The self-oriented subject would be reinforced by effects satisfying his personal needs for domination and ego enhancement regardless of whether his behavior was contributing to the group's effectiveness. The ORI attempts to isolate these types. The subjects choose the most and least preferred of 27 triads of alternatives with test-retest reliabilities of .75 for the three scale scores. Systematic differences in volunteering behavior, in discussion group behavior, and at work have been found, supporting the validity of the ORI (Bass, 1961).<sup>2</sup>

In the present study the subjects used were those 48 of 65 subjects tested who scored in the top quartile of one of the scales but below the median on the other two scales. They were typed according to their highest score (e.g., self-oriented).

#### Procedure

Three groups of subjects were preselected: high scorers on the *Self* scale, high scorers on the *Interaction* scale, and high scorers on the *Task* scale of ORI. Half of the subjects were given agreement (positive reinforcement), the remaining subjects were given disagreement (negative reinforcement) by the partner. On each of five trials the subject discussed the picture with the partner between her initial and final rankings. The measure of the dyadic speech pattern consisted of the duration and frequency of the depressions of the subject's *Talk* and *Listen* keys. Opinion change was measured by the change between the subject's first and final ranking of the faces. In order to standardize the partner's behavior three male graduate students practiced 18 pilot runs which were discarded. Neither the partner nor the experimenter had any knowledge

<sup>2</sup> The test in Bass (1961) includes such items as:

1. One of the greatest satisfactions in life is:
  - A. Recognition of your efforts (S)
  - B. The feeling of a job well done (T)
  - C. The fun of being with friends (I)
2. I think I do my best when:
  - A. I work with a group of people who are congenial (I)
  - B. I have a job that is in my line (T)
  - C. My efforts are rewarded (S)



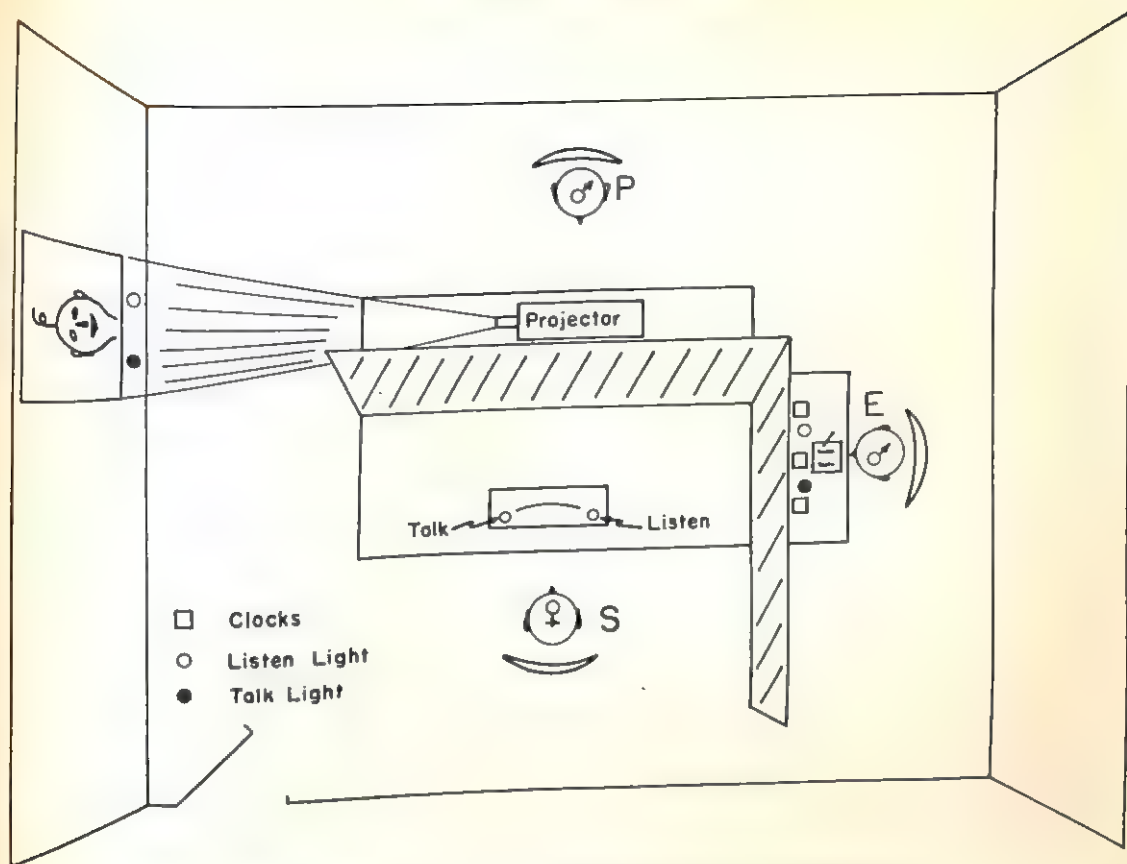


FIG. 1. Schematic diagram of experimental room and apparatus.

of the subject's ORI scores. Assignment to the two experimental treatments was balanced among the partners.

The subject was met in the hall. On the way the experimenter indicated that he had to get another subject first. He said:

We are using two subjects in this experiment and we would prefer that you don't see the other one right now.

When the subject entered the experimental room the partner was seated behind the partition. After the subject was seated, the experimenter asked the subject and the partner for their names and ages individually and presented the instructions. The team was told that the experiment studied the ability to judge personality individually, and in teams. A sample face was projected and the subject and the partner were told to indicate what they thought the face revealed about the personality of the person by ranking their impressions on a pink card. The experimenter continued:

In order to minimize personal biases and to obtain greater objectivity we are also having two people work together as a team. You (pointing to the subject) will be in charge of this team. You can improve the accuracy of your judgment and of your team in the following way. When I say

go ahead, *both* of you should discuss your individual impressions about each picture. You will do this by using these two keys. When you want to talk, press the key marked Talk and keep it down while you are discussing your impressions of the picture. In order to get the help of your partner release your key and press the key marked Listen. Your partner can only talk as long as you press the key. Notice that each key is connected to a set of lights. When you press your key, the red light will go on and you and your partner will know that it is your turn to discuss the picture. When you press your partner's key a green light will go on and your partner will have the opportunity to discuss the picture as long as you keep his key depressed.

The team was told that they would have 4 minutes to discuss each picture. They were asked to rank the picture on the basis of their own impressions and discussion with the partner. The instructions were ended with:

Evaluation of your ability to judge personality accurately will depend on two things, your individual score and also the score obtained by your team. Try to be as accurate as you can, as an individual and as a member of a team.

The experimenter then turned to the partner and said:

Your job differs from that of your partner only in that she is in charge of your team and she will press the appropriate keys. You have no keys and you must take your signal from these lights here.

The experimenter demonstrated the use of the keys by having the subject and the partner interact briefly, and projected the first face. After 1 minute for the initial rankings the experimenter picked up the partner's card (which was actually a blank) and the subject's card, passed the subject's ranking to the partner surreptitiously and signaled the beginning of a trial. At the end of 4 minutes the experimenter asked the team to stop and to rank the traits again on a green card. The same procedure was repeated for five pictures. After the last trial a questionnaire was given to the subject and the partner.

### *Rules for the Partner's Behavior*

Reinforcing statements by the partner were directed only toward the subject's judgment and discussion of the faces. In the agreement group the partner selected from the subject's rankings the traits ranked first, second, third, eighth, ninth, and tenth, and expressed (as if they were his own opinions) statements which generally confirmed the subject's ranking. In the disagreement group the partner expressed opinions which were the reverse of the subject's ratings. The partner avoided answering direct questions and never stated agreement or disagreement directly but only by giving his "own" opinions. The partner continued talking until the subject turned his light off. While his light was on, the partner agreed or disagreed as often as possible. Adherence of the partner to these rules was tested in two ways. The degree of change of the subject's pre- to postratings was subjected to an analysis of variance for the partners under both experimental treatments. No significant  $F$  ratios were obtained, indicating that all partners were equally effective in the extent to which they affected the subject's change in judgments. Secondly, an analysis of variance was performed on the duration of time which the subjects allotted to the three partners. The resultant  $F$  was not significant, indicating that the partners did not differ in the extent to which they influenced the subject's duration of pressing the Listen key.

### *Postexperimental Questionnaire*

The subjects were asked to check scaled answers to the following four questions: How satisfied were you with the group decisions reached? Answers ranged from 1, completely dissatisfied, to 9, completely satisfied. If you were to be re-tested in a team to what extent would you prefer the same partner as you had today? Answers ranged from 1, completely prefer different partner, to 9, completely prefer same partner. To what extent did you participate in discussions in comparison to your partner? Answers ranged from 1, my partner did all of the talking, to 9, I did all of the talking. In comparison to your partner to what extent did you feel influential? Answers ranged from 1, my

partner was completely influential, to 9, I was completely influential.

## RESULTS

### *Opinion Change*

The effects of the experimental treatments and personality variables were examined by an analysis of variance on the rho correlations between each subject's first and second ranking on each trial. A high value of this correlation indicates little shifting of the subject's opinion while a low rho indicates a great degree of change. The analysis yielded a significant  $F$  of 30.81 (1 and 40  $df$ ;  $p < .001$ ) only for agreement versus disagreement. Neither orientation nor the photos affected the extent to which the subject changed her opinion. The mean rho for all subjects under agreement was .30 while the mean rho for subjects under disagreement was .19. The size of the correlations suggests that agreement did not result in very much stability in the subjects' judgments. Considering the ambiguity of the photo judging task, a rho of .30 does indicate that responses to the pictures were not random.

### *Time Allotted to Partner*

The mean duration of Listening and Talking in all groups is presented in Table 1. Table 2 presents an analysis of variance on the total time which the subject allotted to the partner. The subject's willingness to listen to her partner was significantly related to orientation rather than to the reinforcement conditions. Interaction-oriented subjects were most willing to listen, regardless of whether the partner agreed or disagreed (Table 1). An  $F$  of 2.39 for trials (at  $p = .05$ ,  $F = 2.42$ ) revealed that the subjects tended to decrease the amount of time given to the partner as

TABLE 1  
DISTRIBUTION OF TIME (IN SECONDS) SPENT  
IN TALKING AND LISTENING FOR SUBJECTS OF  
VARYING ORIENTATIONS

Orientation	Listen key (second)	% <sup>a</sup>	Talk key (second)	% <sup>a</sup>
Interaction	175.98	73.32	63.32	26.46
Task	165.84	69.10	71.01	29.98
Self	155.50	64.79	83.22	34.86

<sup>a</sup> Percentages do not add to 100, since the remaining time was spent in silence.

TABLE 2  
ANALYSIS OF VARIANCE ON TIME  
ALLOTTED TO PARTNER

Source	df	MS	F
Between subjects	47	1,918.75	
ORI (A)	2	8,384.78	4.93**
Agree versus Disagree (B)	1	44.21	—
A $\times$ B	2	997.80	—
Error	42	1,699.33	
Within subjects			
Trials (C)	4	902.75	2.39*
C $\times$ A	8	328.39	—
C $\times$ B	4	261.81	—
C $\times$ B $\times$ A	8	696.47	1.85
Error	168	376.96	

\*  $p < .10$ .  
\*\*  $p < .025$ .

the experiment progressed. A trend analysis over trials (Edwards, 1960) yielded an  $F$  of 2.61 ( $p < .10$ ) for Orientation and an  $F$  of 3.15 ( $p < .10$ ) for the statistical interaction of Reinforcement and Orientation. In testing the linear slopes of each group against zero it was found that task-oriented subjects who were given agreement significantly decreased in time allotted to their partner over trials ( $p < .05$ ). Interaction-oriented subjects when given agreement also tended to decrease ( $p < .10$ ), while self-oriented subjects under agreement tended to increase ( $p < .10$ ). All groups under disagreement showed slopes which did not differ from zero. These results suggest that the change in time allotted to a partner was a function of the subject's orientation as well as reinforcement conditions.

#### Frequency of Interactions

When frequency of presses for the Listen key was subjected to an analysis of variance, neither reinforcement conditions nor Orientation effects were found to be significant. An  $F$  of 2.20 with 4 and 168  $df$  ( $p < .10$ ) was found for trials, suggesting a decrease of initiating interactions over trials. These results were examined further by a trend analysis. The linear components in each group did not differ significantly over trials. When the quadratic components were examined, the reinforcement conditions differed at  $p < .10$  ( $F = 3.25$ ) and the Reinforcement by Orientation

interaction was significant at  $p < .025$  ( $F = 4.01$ ). With agreement, the groups tended to decrease rapidly over the first three trials and then increased until they reached the same frequency on the fifth trial as on the first trial. On the other hand, the groups receiving disagreement tended to listen with approximately the same frequency over all trials. In this regard, the interaction-oriented subjects also showed the least change over trials, while the task-oriented subjects differed most, yielding a pattern of initial decrease followed by recovery. Self-oriented subjects showed this pattern to a lesser degree.

#### Subject Talk-Time

Since the subjects were requested to depress either of the two keys throughout each trial, there is an inverse relationship between the amount of time allotted to the partner and the amount of time taken by the subject. However, since some time (actually less than 1% of total time available) could be taken up by silence (failure to depress either button), an additional analysis was performed on the amount of time during which the subject depressed his Talk key (see Table 1). An analysis of variance yielded an  $F$  of 4.81 ( $p < .025$ ) for Orientation. The analysis further indicated a significant  $F$  of 3.27 (with 4 and 168  $df$ ,  $p < .025$ ) for trials. All subjects tended to talk more as the experiment progressed. A trend analysis substantiated the findings from the analysis of time allotted to the partner. Interaction and task-oriented subjects under agreement significantly increased in time spent talking on successive trials ( $p < .05$ ).

The analysis of the subject's own speech patterns indicated again that orientation was relatively more important in determining the subject's dyadic speech pattern. Self-oriented subjects took the most time, interaction-oriented subjects spent the least time talking. All subjects invited the partner to talk more than they talked themselves (see Table 1). The mean percentage for the agreement groups was 30.9%; for the disagreement groups it was 30.0%.

An analysis of variance on the frequency with which the subject pressed the Talk key yielded no significant differences with regard



to the main variables. An  $F$  of 2.23 with 4 and 168  $df$  ( $p < .10$ ) was obtained for trials.

### *Questionnaire Data*

Each of the four postexperimental questions was analyzed by an  $F$  test for differences between reinforcement conditions and Orientation scores. The subjects who were given agreement yielded a mean score of 7.67 in response to question one, "How satisfied were you with the group decisions reached?" while subjects given disagreement yielded a mean score of 4.29 ( $F = 79.94$ ,  $p < .001$ ). On question two, the subjects given agreement had a mean score of 7.17 in how much they preferred their partner for a retest, while the subjects given disagreement had a mean score of 4.79 ( $F = 26.33$ ,  $p < .001$ ). On question three, about how much participation was shared by the subject and the partner, significant differences were found for orientation but not for reinforcement conditions ( $F = 8.33$ ,  $p < .001$ ). Self-oriented subjects stated that they talked as much as their partner ( $M = 4.69$ ), task-oriented subjects felt their partner talked more ( $M = 4.06$ ), and interaction-oriented subjects felt their partner talked much more ( $M = 3.69$ ).

As can be seen from Table 1, these ratings accurately reflected the talk-listen differences between the subject and the partner. However, all subjects grossly overestimated their own contributions in the interaction. On question four, about how influence was shared, both reinforcement conditions and orientation yielded significant differences. An  $F$  of 8.75 ( $p < .005$ ) indicated that the subjects under agreement felt they had equal influence in making the decision ( $M = 5.33$ ) while the subjects with whom the partner disagreed thought the partner was more influential ( $M = 4.38$ ). The analysis yielded an  $F$  of 3.59 ( $p < .05$ ) for orientation. The interaction-oriented subjects felt that their partner was more influential ( $M = 4.25$ ), and the task-oriented subjects felt that they were of about equal influence ( $M = 5.06$ ). Self-oriented subjects thought that they were somewhat more influential than their partner ( $M = 5.25$ ).

### DISCUSSION

The present findings yield evidence that the content of the partner's speech affected

both the subject's opinions and the time which she allotted to herself and to her partner. The subjects shifted opinions more when interacting with a disagreeing partner than an agreeing partner. In this regard, the subjects' orientation was not a relevant factor. The correlation coefficients indicated that the difference in opinion change was not very great, even though it was significantly different for the two experimental treatments. However, in the present study the variables which usually result in strong modification of group opinion have been minimized. For example, the subject never saw her partner. She considered him as a peer at best, since "the subject had been put in charge of the team" by instructions. The partner played a subservient role and avoided expression of strong opinions. The results indicate that even under these conditions disagreement enhanced the subject's flexibility.

The subjects' tolerance for listening to their partners decreased and their willingness to talk increased with time. This tendency may have been due to boredom with the partner, or a gradual lowering of the estimation of the partner's importance. Or, the subjects may have become more comfortable in the interaction, thus tending toward a more equal distribution of time. However, even during the last trial the subjects talked for much less than half the available time. These results have some implication for the behavior of the counselor in therapeutic interactions. If tolerance for listening to another person decreases with increased exposure, then a therapist would have to guard against a tendency to increase talking as a counseling session progresses.

The personality traits, measured by the ORI, strongly influenced the subject's tendency to approach, listen to, or talk to the partner. Given control over the interaction, subjects allotted over two-thirds of the total available time to their partners. The self-oriented subjects behaved as reinforcement theory would suggest all subjects should. As they experienced agreement from their partners they increased the time allotted to listening, thereby increasing the number of positive reinforcements for their opinions.

On the other hand, interaction and task-oriented subjects behaved more like a dynamic theory would suggest all subjects should. As they experienced agreement from the partner their attempts to influence him increased and the time allotted him decreased. Apparently, self-oriented subjects attempted to directly magnify the partner's approving statements by inviting him to talk for longer periods. This involved no effort since the subjects only had to press the Listen key in order to obtain agreement. On the other hand, interaction and task-oriented subjects increased the behavior for which they had been rewarded. These subjects focused on their own talk, while the self-oriented subjects focused on partner's agreements. Disagreement produced no significant change in the subject's willingness to listen to the partner's opinions.

Combining both experimental conditions it was found that interaction-oriented subjects were most willing to listen to their partners. Presumably these subjects may see their willingness to listen as a means of maintaining good relations with him. Self-oriented subjects appeared least concerned about listening, possibly showing their greater independence in this situation. When the self-oriented subject is in a free small group discussion his behavior is quite different. He tends to talk the least and to successfully influence the group very little (Bass, 1961).

Since all subjects were female and the partners were male the sex composition of the dyadic team may have been a factor affecting the distribution of time among participants. In a recent study by Elwood (1963), using a similar procedure, both male and female subjects were used with male partners. Elwood found higher mean durations for male subjects than for female subjects under some of his conditions, while sex was not an effective variable in other conditions. A comparison of the present findings with Elwood's results further suggests that the task variable also affects the distribution of talk-time. Given no specific task, the subjects were simply instructed to engage in conversation to "get better acquainted." Under these conditions the subjects also talked less at first but increased over trials. The final percentage of time allotted to the

partner closely approximated 50%. From the results of both studies it is evident that the status of the partner, the subject's competence in solving a problem, and the sex relationship in the team all affect dyadic speech patterns. On the questionnaire the subjects felt more satisfied with the group discussion when the partner agreed than when he disagreed. The agreeing partner was also much preferred and the subjects felt that he had more influence under agreement than disagreement. These results suggest the importance of the personal and emotional by-products affecting the quality of the dyadic relationship, quite apart from the real value of a partner's contribution. These data lead one to suspect that conformity of a partner may become a more important criterion for his inclusion in a dyad or group than his actual contribution to a task. Social orientation also affected the subject's impression about his dyadic role. Self-oriented subjects tended to overrate their own importance in the interaction more, perceived themselves as more influential and more actively participating (as indeed they were) than subjects who were task or interaction-oriented.

#### REFERENCES

- BASS, B. M. *Leadership, psychology, and organizational behavior*. New York: Harper, 1960.
- BASS, B. M. Comparisons of the behavior in groups of self-oriented, interaction-oriented and task-oriented members. Technical Report No. 25, 1961, Louisiana State University, Contract N7, ONR 35609.
- BASS, B. M. *The Orientation Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- CAMPBELL, O. H. Objective behavior in dyads of self, interaction and task-oriented members. Unpublished master's thesis, Louisiana State University, Baton Rouge, 1961.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960.
- ELWOOD, D. L. Modification of manding in dyadic interactions. Unpublished doctoral dissertation, Purdue University, Lafayette, 1963.
- KANFER, F. H., & McBREARTY, J. F. Minimal social reinforcement and interview content. *J. clin. Psychol.*, 1962, 18, 210-215.
- SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- STRITCH, T. M. Experimentally induced changes in physiognomy and personality impressions. Unpublished doctoral dissertation, Emory University, Emory, 1954.

(Received February 12, 1962)



# QUESTIONNAIRE MEASURE OF THE HYSTEROID/OBSESSOID COMPONENT OF PERSONALITY:

## THE HOQ

T. M. CAINE

*Claybury Hospital, Woodford Green, Essex, England*

AND L. G. HAWKINS<sup>1</sup>

*Central Office of Information, London, England*

In line with recent studies distinguishing between symptom and personality trait complexes a questionnaire purporting to measure the hysteroid/obsessoid dimension of personality was devised. The questionnaire was administered to 93 neurotic patients in a mental hospital who were rated by the staff on the relevant personality traits. An analysis of the ratings was carried out and the questionnaire was related to the rating scale classification and to other clinical and psychological test material as available. The questionnaire was found to relate to personality measures more closely than to symptom measures with a correlation of .68 with the rating classification significant at the .001 level. Arguments were advanced as to why the questionnaire should be regarded as primarily a measure of the hysteroid/obsessoid dimension.

MacKinnon (1944) has credited Janet with the first attempt to distinguish between symptom and personality trait constellations in the neuroses. A revival of interest in this problem has taken place in recent years as attempts have been made to improve the validity and reliability of psychiatric diagnoses and to devise relevant measuring instruments (Chodoff & Lyons, 1958; Foulds, 1961b; Foulds & Caine, 1958, 1959; Sandler & Hazari, 1960). The distinction is of relevance to therapists since there is evidence to suggest that psychological tests related to symptomatology change more readily with treatment than do tests related to personality trait constellations, with the personality trait measure here to be described remaining relatively more stable than symptom and attitude tests (Foulds, 1959; Foulds & Adams, 1962; Martin & Caine, 1963).

The recent empirical evidence thus supports Janet in his contention that the hysteroid and obsessoid personalities can be usefully dichotomized along a single dimension as an examination of the traits making up the relevant constellations suggests. Until the development of the Hysteroid/Obsessoid

Questionnaire (HOQ), no test based directly upon the traits involved and validated against an outside criterion has been developed.

The present study is a description of the construction, validity and reliability of the HOQ together with its relation to other relevant factors. If the HOQ is found to relate more closely to hysteroid/obsessoid trait constellations than to Hysteria/Dysthymia symptom complexes, further justification for the separation of symptoms from traits will have been provided.

## METHOD

### *Construction of the Questionnaire*

It was found in previous work that self-ratings by hysteroid and obsessoid persons on the hysteroid-obsessoid rating scale described by Foulds and Caine (1958) failed to differentiate the two groups sufficiently accurately to be of much practical value. It was felt that rating scale principles were difficult to grasp for untrained subjects, that some of the points had moral implications as worded, and that the 9-point scale employed was too short for high reliability. To escape these difficulties the HOQ was constructed by framing a number of statements thought to be pertinent to each of the nine rating scale traits. Rating scale item "desire to impress and gain attention—self-effacing," for example, formed the basis of the following four true-false statements:

<sup>1</sup> Thanks are due to Keith Hope for his help with some of the statistics.



- I like to wear eye-catching clothes
- I keep quiet at parties or meetings
- It pleases me to be the center of a lively group
- I like discussing myself with other people

Forty-eight statements were thus compiled each of which could be scored in a hysteroid or obsessoid direction.

### *Administration of Tests*

This questionnaire was administered to all patients entering the community wards for neurotics at Claybury for an 18-month period. In addition all neurotic patients referred to the psychologist from the acute admission wards during the period as hysteroid or obsessoid personalities were tested. A minimum of Grade IV above the tenth centile on the Mill Hill Vocabulary Scale was set. Ninety-three patients in all were tested on the HOQ and of these 77 were given the Minnesota Multiphasic Personality Inventory (MMPI) scales found to distinguish Hysterics and Dysthymics in the Foulds and Caine (1958, 1959) studies. These were combinations of the *Hs*, *D*, and *Pt* scales with the *D-Hs* score for women and the *D+Pt* score for men. *K* was also included as were the Extra-punitive and Intropunitive scales reported in detail by Foulds, Caine, and Creasey (1960). Fifty-three of these patients also completed the Maudsley Personality Inventory (MPI) but at different stages of treatment, together with the HOQ. Sixty-two were retested on the HOQ 6 weeks after admission.

### *Establishment of the Criterion*

Ratings on all patients were obtained from medical, nursing, and occupational therapy staff, after training, using a form of the rating scale described in the previous study (Foulds & Caine, 1958). The number of raters per patient varied from 3 to 11 with a mean of 7.5. Ratings were usually completed towards the end of the second week after admission, when raters felt reasonably confident of making an accurate assessment.

An analysis of the internal consistency of the rating scale items showed that only 6 of the 11 original traits were being successfully differentiated for the groups by the raters, namely, attention seeking, emotional display, speed of decision, lability of affect, conscientiousness, and shallowness of affect. These six items formed the subsequent basis of the hysteroid-obsessoid rating allocation.

An analysis of the mean rating was computed and from this confidence limits were calculated. Despite the safeguards taken in the rating scheme these calculations showed that the dependability of the ratings was in question. Although this is a common and recognized weakness of ratings, a factorial experiment was carried out in an effort to improve the validation criterion. Five of the raters each rated 10 males and 10 female patients twice, with a 10-day interval between ratings, making a total of 200 ratings. An analysis of variance was carried out with raters, replica-

tions of ratings, and patients as the sources of variance. The between-raters, between-patients, and rater/patient interaction variances were significant but the between replications were not. The between raters variance was too small to be of practical significance and since the rater/patient interaction could not, in practice, be allowed for, it was included in the residual variance. The experiment was collapsed into a 5 (raters)  $\times$  20 (patients) factorial design. Estimates of the variance of the factors were made on the assumption of Model 11 of the analysis of variance (Snedecor, 1946). The variance of the raters was only .35, while that of the patients was 4.37. The error variance was 2.69. One may conclude from the factorial experiment that raters tended to show something of a bias in one direction or another but that this bias was too small to be worth correction, and that the raters differed frequently from each other in their individual classifications. The reliability of the mean rating of each patient (averaged over the five raters and measured by an intraclass correlation coefficient) was .92.

The distribution of the patients' ratings was not significantly different from a normal distribution having the same mean and standard deviation. Degree of hysteroidness or obsessoidness as observed by trained observers was thus not related to hospital admission.

To establish the hysteroid or obsessoid rating classification the ratings for each patient were pooled, taking the total hysteroid score less the total obsessoid score as the allocation. This allocation yielded 65 obsessoids, 27 hysteroids, and 1 indeterminate.

A preliminary analysis of the HOQ was conducted to determine the score association of questionnaire items among themselves. Items most closely related were placed together with agreements below the .05 level of significance being ignored. A factor analysis of the questionnaire is contemplated when a larger number of cases has been collected.

The score on the HOQ (total hysteroid score since each item could be scored hysteroid or obsessoid) and other clinical and test data were related to the personality rating classification and to themselves.

## RESULTS

### *Abilities, Sex, and Age Composition of the Sample*

Of the 93 cases (42 men, 51 women) 91 completed the Mill Hill Vocabulary synonym test. The mean synonym score of these was 27, giving an expected total of 51, which is well within the "average" range. Of the two remaining referred cases, one had a Wechsler IQ of 105, and the other fell at the seventy-fifth centile on the Progressive Matrices test. The mean age of the 28 rated

hysteroids was 33.72 (*SD* 8.78), that of the 65 rated obsessoids was 33.95 (*SD* 9.50) with a directly comparable age range of 19-54.

### *Predictive Power of the Questionnaire*

The power of the questionnaire in excess of chance in predicting the rating scale allocation was calculated taking a score of 24 and above as the questionnaire hysteroid classification. To call a person hysteroid or obsessoid is similar to calling him long or short—the dividing line is arbitrary. A priori, since only hysteroid responses are scored on the HOQ, those scoring on more than half the items may be legitimately assigned to the questionnaire hysteroid classification. The rating scale classification was thus found to be correctly predicted by the questionnaire in 80% of the cases, an improvement over chance of 26%.

### *Correlations with the HOQ*

Biserial, product-moment, and Pearson *r*'s were calculated between the HOQ and other clinical data and tests and the results are shown in Table 1.

No significant relationship was found between any of the clinical scales of the MMPI and the rating scale classification. There was no significant sex difference in HOQ score of those rated hysteroid or obsessoid although significantly more women were rated as hysteroid. Scores on the HOQ were normally distributed.

### *Item Agreement within the HOQ*

Items showing the highest score agreement within the questionnaire could be identified under the following traits: (a) imaginative thinking; (b) meticulous concern with details and standards; (c) self-effacing; (d) self-display, fashionable dress; (e) speed of decision; (f) conscientiousness; (g) deep, lasting friendships; (h) dependence on others; (i) inhibition of emotions; (j) lability of effect. All items tended to agree with the exception of one to which 83 subjects responded "yes" and which was misplaced in the item agreement analysis.

### DISCUSSION

Table 1 shows that with the exception of the MPI *E* Scale the highest measure of

TABLE 1

CLINICAL AND TEST DATA CORRELATIONS WITH THE  
HYSTEROID-OBSOSSOID QUESTIONNAIRE

Data	Number of cases	Correlation coefficient
HOQ retest (after 6 weeks community therapy)	62	.77***
Rating classification	92	.68***
Clinical diagnosis (Hysterics 15; Dysthymics 71)	86	.28**
Age	93	.23*
Verbal ability (Mill Hill Vocabulary scale)	91	-.08
MMPI scales		
<i>K</i>	77	.08
<i>Hs</i>	77	-.14
<i>D</i>	77	-.15
<i>Pt</i>	77	-.11
Test diagnosis (Hysteria; Dysthymia)		
<i>D - Hs</i> (women)	41	.002
<i>D + Pt</i> (men)	36	.32
Total punitive score		
MMPI Extrapunitive + Intrapunitive	77	-.07
Direction of hostility		
MMPI Extrapunitive - Intrapunitive	77	.31**
Maudsley Personality Inventory (MPI)		
Extraversion	53	.84***
Neuroticism	53	-.40***
Correlation between MPI Extraversion and Neuroticism		-.46

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

agreement is between the HOQ and the personality rating classification. There is a considerably smaller measure of agreement between the HOQ and the clinical diagnosis. The test diagnosis for men is almost significant at the .05 level but that for women is negligible. The total punitive score is not related to personality as measured by the HOQ but hysteroids are significantly more extrapunitive, a finding supportive of previous work. There is a significant positive



relationship between age and HOQ in this Claybury sample.

The very high level of agreement between the HOQ and the MPI *E* scale suggests that they are measuring virtually the same thing. This is not surprising since Eysenck (1959) places more reliance on "attitudes" than on symptoms in hysterical symptom questionnaire construction and equates hysterical conversion symptoms, "hysterical attitudes," hysterical personality, and psychopathy in establishing his extraversion criterion groups. He also points to the heavy dependence of the *E* Scale on "sociability" (Eysenck & Claridge, 1962), a trait which is related to some of the self-effacing versus attention seeking items of the HOQ.

It might well be that the failure to distinguish between trait and symptom complexes has led to the lack of confirmation of Eysenck's initial findings (Eysenck & Claridge, 1962; McGuire, 1962; Sigal, Star, & Franks, 1958). As far as the author is aware no attempt has been made to validate the MPI *E* Scale against an outside criterion of extraversion as measured by ratings of associated personality traits. The logical fallacies of using Hysterics and Dysthymics for this purpose have been pointed out by Foulds (1961a), and the present study supports his contentions. In view of the theoretical controversy and conflicting experimental evidence concerning Eysenck's system, one is justified in restricting the interpretation of the HOQ to the hysteroid-obsessoid dichotomy for which it was devised. The introduction of a third extraneous and controversial concept is not required in the interpretation of the data. Particularly is this so when there is general agreement in the literature that the two personality types in question form recognizable clinical entities.

## REFERENCES

- CHODOFF, P., & LYONS, H. Hysteria, the hysterical personality and "hysterical" conversion. *Amer. J. Psychiat.*, 1958, 114, 734-740.
- EYSENCK, H. J. *The structure of human personality*. (2nd ed.), London: Methuen, 1959.
- EYSENCK, H. J., & CLARIDGE, G. The position of hysterics and dysthymics in a two-dimensional framework of personality description. *J. abnorm. soc. Psychol.*, 1962, 64, 45-46.
- FOULDS, G. A. The relative stability of personality measures compared with diagnostic measures. *J. ment. Sci.*, 1959, 105, 783-787.
- FOULDS, G. A. The logical impossibility of using hysterics and dysthymics as criterion groups in the study of introversion and extraversion. *Brit. J. Psychol.*, 1961, 52, 385-387. (a)
- FOULDS, G. A. Personality traits and neurotic symptoms and signs. *Brit. J. med. Psychol.*, 1961, 37, 1-8. (b)
- FOULDS, G. A., & ADAMS, A. Depression and personality. *J. ment. Sci.*, 1962, 108, 474-486.
- FOULDS, G. A., & CAINE, T. M. Psychoneurotic symptom clusters, trait clusters and psychological tests. *J. ment. Sci.*, 1958, 104, 722-731.
- FOULDS, G. A., & CAINE, T. M. Symptom clusters and personality types among psychoneurotic men compared with women. *J. ment. Sci.*, 1959, 105, 469-475.
- FOULDS, G. A., CAINE, T. M., & CREASY, M. A. Aspects of extra- and intro-punitive expression in mental illness. *J. ment. Sci.*, 1960, 106, 599-610.
- MCGUIRE, R. J. A study of the M. P. I. used with psychiatric in-patients. *Bull. Brit. Psychol. Soc.*, 1962, 47, 56-57.
- MACKINNON, D. W. The structure of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 1. New York: Ronald Press, 1944.
- MARTIN, D. V., & CAINE, T. M. Personality change in chronic neurosis in a therapeutic community. *Brit. J. Psychiat.*, 1963, 109, 267-272.
- SANDLER, J., & HAZARI, A. The "obsessional": On the classification of obsessional character traits and symptoms. *Brit. J. med. Psychol.*, 1960, 33, 113-122.
- SIGAL, J. J., STAR, K. K., & FRANKS, C. M. *J. abnorm. soc. Psychol.*, 1958, 57, 143-148.
- SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Iowa State Coll. Press, 1946.

(Received July 2, 1962)



## DEVIANT RESPONSE PATTERNS AS A FUNCTION OF CHRONOLOGICAL AGE<sup>1</sup>

S. O. HESTERLY<sup>2</sup>

*Louisiana State University*

Berg's Deviation Hypothesis suggested that: (a) response patterns of children on the Perceptual Reaction Test (PRT) will vary with age to adulthood so that PRT age scales can be developed reflecting maturity level; (b) PRT age scale scores of elderly adults will reflect a decline in maturity level. PRT responses which differentiated 500 children aged 6 from 850 adults aged 20-59 ( $p < .0025$ ) were selected for the age scales. PRT age scale scores for cross-validation groups (500 children aged 7-16; 203 adults aged 20-59; 200 adults aged 60-83) reflected systematic increases in scores with age to adulthood ( $p < .001$ ) and decreases in scores for elderly adults ( $p < .001$ ). The scores of elderly adults approximated those of children aged 9-10.

Subjects seldom respond randomly to multiple-choice tests even when there is no obvious basis for choosing one response above another. Instead, certain responses are chosen significantly more often than others, so that biases occur. Berg, Hunt, and Barnes (1949) developed a test which was specifically designed to elicit such biases. This test, the Perceptual Reaction Test, referred to hereafter as the PRT, consists of 60 abstract designs to which the subject responds by choosing one of the following options: like much, like slightly, dislike slightly, or dislike much. Of course, the subject can make no response, although this is contrary to the test instructions. When Berg (1955, 1957) administered the PRT to different groups of subjects, he found that subjects who made responses which departed from the typical bias patterns, i.e., uncommon or atypical responses, also tended to show personality deviations. These findings led him to formulate the Deviation Hypothesis which states (Berg, 1957):

Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality (atypicalness) and thus regarded as symptoms (earmarks or signs) are asso-

ciated with other deviant response patterns which are in noncritical areas of behavior and which are not regarded as symptoms of personality aberration (nor as symptoms, signs, earmarks) [p. 159].

A series of studies (Adams, 1959; Barnes, 1955; Grigg & Thorpe, 1960; Harris, 1958; Hesterly & Berg, 1958; Roitzsch & Berg, 1959) were stimulated by the early studies. The Hesterly and Berg (1958) study provided the suggestion for the present research. The results obtained in this investigation indicated that the response patterns of young children to the PRT differed significantly from those of normal adults in such a way as to approximate the deviant patterns which Barnes (1955) had found with schizophrenic adults. In addition, there was clearly a significant tendency for scores on Barnes' schizophrenia scale to decrease with age, so that older children tended to score more like normal adults.

The present study represents an attempt at using these changes with age in the response patterns of normal children as a measure of maturity as defined by chronological age. The Deviation Hypothesis (Berg, 1957) suggests that the various "critical" behavior patterns which are associated with the physical and psychological changes which occur as children grow to maturity should be reflected in "noncritical" response patterns on the PRT, which change with increasing age. Specifically, an attempt was made to develop age scales for the PRT, using the normal adult data employed by Barnes (1955) as

<sup>1</sup> This article is based on a dissertation submitted to the Graduate School, Louisiana State University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is grateful to his advisor, I. A. Berg, for his aid and encouragement.

<sup>2</sup> Now at the University of Arkansas, School of Medicine, Little Rock, Arkansas.

a criterion measure. Further, the possibility that normal adults, as they approach senility, again manifest deviant patterns on the PRT, similar to those of children, was also investigated. Accordingly, the hypotheses of the present study were two:

1. The response patterns of normal children on the PRT will vary with age in such a way as to reflect maturity level as defined by chronological age, making possible the development of age scales based on group tendencies.

2. The response patterns of normal adults on the PRT will vary with age, when older age groups are considered, in such a way as to reflect decreases in scores on the age scales as they approach senility.

In short, a systematic development and decline in patterns of response on the PRT was hypothesized for children and aged persons. Normal children and elderly adults would be expected to manifest similar patterns of deviant responses to the PRT. This represents a test of the Deviation Hypothesis (Berg, 1957). The behaviors shared by normal children and elderly adults may be regarded as critically different from normal adults aged 20 through 59 years. Such critical differences should be reflected in non-critical areas of behavior as PRT option selections.

## METHOD

### Subjects

A total of 2,253 subjects were used in the present study. Five hundred school children (divided equally by sex) from the public and parochial schools of two Louisiana parishes aged 6-0 through 6-11 served as subjects in the construction of the PRT age scales, with 850 normal adults (500 males; 350 females) from Barnes (1955) constituting the normal adult sample.

Five hundred public school children, from four Louisiana parishes, divided equally by age and sex, were used in developing norms for older children and in testing the validity of the PRT age scales at each of the following age levels: 7-0 through 8-11; 9-0 through 10-11; 11-0 through 12-11; 13-0 through 14-11; 15-0 through 16-11.

Two hundred three normal adults (105 males; 98 females) drawn from Berg's file of PRT records served as a sample of normal adults aged 20 through 59 years. This sample included policemen, school teachers, housewives, adults in evening college classes, and the like. Two hundred elderly adults (50 males; 150 females) ranging in age from 60

to 83 years, formed the sample used in testing the second hypothesis of this study and in developing norms for elderly people. These subjects were all members of the Golden Agers Organization of Orleans Parish, Louisiana and included individuals from various socioeconomic backgrounds.

### Procedure

The PRT was administered to the subjects in groups of 12 to 70. Standard instructions were used with adults; however, these instructions were modified somewhat, as in a previous study (Hesterly & Berg, 1958), for children, to insure their understanding the nature of the task required of them. The responses for each subject were transferred to IBM answer sheets for tabulation and scoring. For each option, the percentage of the criterion (or normal adult) group was compared with the percentage of the experimental (or normal 6-year-old) group and tested for statistical significance by use of a contingency table (Mainland & Murray, 1952). Age scales for males and females, taken separately, were then constructed by a cross-validation technique proposed by Katzell (1951). In this technique, the samples are divided into random halves with independent statistical analyses performed on each half and the resulting scales validated on the remaining half. Only those options which differentiated the groups beyond the .05 level of confidence on both scales were included in the final scales. Each option preferred by normal adults was given a weight of 1; each option preferred by normal 6-year-old children was given a weight of -1. Each subject's PRT age scale score was computed by algebraic summation of the weights for the options he chose. Measures of central tendency, variability, and reliability were then computed for the various age groups. The variances for the groups appeared to be rather heterogeneous; therefore Bartlett's test for homogeneity of variance was applied to the five age groups of children and Fisher's test for homogeneity of variance between two groups was applied to children versus adults, children versus elderly adults, and adults versus elderly adults. Since all of these tests yielded probabilities of less than .05, nonparametric techniques were used to compare the group PRT age scale scores. Specifically, the nonparametric median test was deemed appropriate for data of the type obtained, as this test is not dependent upon homogeneity of variance. The hypothesis tested is that the groups compared are random samples from a population with a common median.

## RESULTS AND DISCUSSION

The male PRT age scale constructed in this study included 176 of the 300 possible response options. Sixty-four options were significantly more frequent among 6-year-olds, and 112 options among adult males; accordingly, weights of +1 and -1, respectively,



TABLE 1  
SUMMARY OF MEASURES OF CENTRAL TENDENCY, RELIABILITY, AND VARIABILITY OF  
PRT AGE SCORES BY AGE GROUPINGS

Age group	Males					Females				
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>r<sub>tt</sub></i>	<i>SE<sub>est</sub></i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>r<sub>tt</sub></i>	<i>SE<sub>est</sub></i>
6 to 6-11	250	-21.0	23.80	.97	4.04	250	-15.6	18.30	.96	3.66
7 to 8-11	50	-12.5	30.64	.98	4.29	50	-10.3	19.52	.98	2.76
9 to 10-11	50	-0.3	23.39	.95	5.23	50	0.4	16.45	.99	1.64
11 to 12-11	50	5.5	13.09	.99	1.31	50	5.3	15.68	.99	1.57
13 to 14-11	50	13.9	18.55	.99	1.86	50	11.0	15.71	.98	2.22
15 to 16-11	50	23.1	16.85	.97	2.92	50	18.2	10.68	.90	3.37
7 to 16-11	250	5.9	23.96	.99	2.40	250	5.6	18.02	.99	1.80
20 to 59	105	29.1	15.29	.95	3.42	98	22.4	12.86	.93	3.40
60 to 83	50	2.1	27.15	.99	2.72	150	-4.1	19.93	.91	6.00

were assigned to each option. Thirty-seven other options were eliminated in the cross-validation.

Similarly, the female PRT age scale included 135 options (55 significantly more frequent among 6-year-olds, 80 among adult females). Forty-three other options were eliminated in the cross-validation.<sup>3</sup>

A summary of measures of central tendency, variability, and reliability for various age groups on the PRT age scales is presented in Table 1. A systematic negatively accelerated function in mean score as a function of age from age 7 through age 16 can be observed.

The results shown in Table 2 indicate that a statistically significant increase in median PRT age scale score occurs among normal children aged 7-0 through 16-11. With increasing age, these scores increasingly resemble those of adults aged 20 through 59. This suggests that the scales would be correlated with the critical behavioral changes which occur during the process of growing to adulthood, e.g., those response modifications which are related to physical and intellectual maturation.

The results shown in Table 3 indicate that the PRT age scales significantly differentiated children from adults aged 20 through 59. Further, the female PRT age scale sig-

nificantly differentiated children from elderly adults, with the latter obtaining a significantly lower median score. A similar but nonsignificant trend was present with the male PRT age scale. Inspection of the mean and median scores reveals that the elderly adults in the present study responded to the PRT in a manner most like that of children aged 9-0 through 10-11. The Deviation Hypothesis (Berg, 1957) would lead us to infer from these results that elderly adults and children aged 9-0 through 10-11 share common behavior patterns in a critical area.

Table 3 also indicates significant differences between the PRT age scale scores of adults aged 20 through 59 and elderly adults, with the latter obtaining significantly lower median scores. From this it may be inferred that a decline in PRT age scale scores exists which is associated with the critical behavioral changes occurring during later adulthood, e.g., those response modifications related to physical and intellectual decline.

Thus it would seem that the PRT age scales measure a function which resembles the general pattern of growth and decline which has previously been observed in manual skills (Miles, 1931), strength (Cathcart, Hughes, & Chalmers, 1935; Ruger & Stoessiger, 1927), reaction time (De Silva & 1936; Pearson, 1925), intelligence (Jones & Conrad, 1933; Miles, 1931), and learning ability (Miles, 1933; Ruch, 1934). However, the task of responding to the designs of the PRT seems to be qualitatively different from

<sup>3</sup> The complete dissertation containing keys and *T* scale conversion tables for the PRT age scales may be obtained by ordering Publication No. 60-5916 from University Microfilms, Ann Arbor, Michigan.



TABLE 2

NONPARAMETRIC MEDIAN TEST FOR PRT AGE SCORE DIFFERENCES BETWEEN AGE GROUPS OF CHILDREN

Age groups	Below median	Above median	Subgroup median	Group median	df	$\chi^2$
Males						
7-0 to 8-11	38	12	-8.0	7.9	4	43.63**
9-0 to 10-11	28	22	4.0			
11-0 to 12-11	25	25	7.5			
13-0 to 14-11	18	32	11.6			
15-0 to 16-11	7	43	22.7			
Females						
7-0 to 8-11	38	12	-7.6	5.7	4	227.85**
9-0 to 10-11	32	18	1.6			
11-0 to 12-11	22	22	3.6			
13-0 to 14-11	21	29	10.5			
15-0 to 16-11	5	45	17.4			

\*\*  $p < .001$ .

those involved in the studies above in two ways. The PRT instructions suggest the absence of "correct" or desired responses. Hence, the subject is asked to make a choice without the implication that he should attempt to choose correctly or to display a skill, strength, etc. In addition, the response categories (like much, like slightly, etc.) imply that the subject is expected to give an

affective response. These aspects of the PRT task suggest that performance with the PRT is more likely to reflect differences in affective reactivity than differences in skill, intelligence, strength, and the like.

With these considerations in mind, it is possible to infer that the PRT age scales tap some kind of affective maturity which has a systematic development and decline

TABLE 3

NONPARAMETRIC MEDIAN TEST FOR PRT AGE SCORE DIFFERENCES BETWEEN PAIRED AGE GROUPS

Age groups	Below median	Above median	Subgroups median	Group median	df	$\chi^2$
Males						
7-0 to 16-11	165	85	7.9	15.40	1	78.95**
20 to 59	15	90	34.2	7.10	1	2.21
7-0 to 16-11	121	129	7.9			
60 to 83	30	20	4.5	27.50	1	25.89**
20 to 59	36	69	34.2			
60 to 83	39	11	4.5			
Females						
7-0 to 16-11	150	100	5.7	11.98	1	42.92**
20 to 59	21	77	24.4	3.50	1	9.76*
7-0 to 16-11	108	142	5.7			
60 to 83	39	61	-0.9	9.00	1	83.62**
20 to 59	115	83	24.4			
60 to 83	112	38	-0.9			

\*  $p < .01$ .\*\*  $p < .001$ .

with age in normal individuals. The data tend to support this inference in that the normal adults aged 20 through 59 generally chose the "like slightly" or "dislike slightly" options for the vast majority of their responses. This suggests that their affective reactions to the designs are rarely extreme. On the other hand, children and elderly adults chose the "like much" option quite frequently, suggesting relatively more extreme affective reactions to the designs. Perhaps these test performances reflect the characteristics which children and elderly people tend to share, viz., immaturity, childishness, atypical affect. The results of Barnes (1955) indicate that schizophrenic patients also tend to choose the "like much" option for many of their responses. This finding also lends support to the inference that performance with the PRT reflects affective maturity, since schizophrenics are typically characterized as immature, childish, and atypical in affect. The author is currently engaged in further research with the PRT age scales in an attempt at assessing the inference.

## REFERENCES

- ADAMS, H. E. Deviant responses to auditory and visual stimulus patterns in the identification of schizophrenia. Unpublished master's thesis, Louisiana State University, 1959.
- BARNES, E. H. The relationship of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, 51, 286-290.
- BERG, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, 40, 61-72.
- BERG, I. A. Deviant responses and deviant people: The formulation of the deviation hypothesis. *J. counsel. Psychol.*, 1957, 4, 154-161.
- BERG, I. A., HUNT, W. A., & BARNES, E. H. *The Perceptual Reaction Test*. Evanston, Ill.: IAB, 1949.
- CATHCART, E. P., HUGHES, D. E. R., & CHALMERS, J. G. The physique of men in industry. Report No. 71, 1935, Medical Research Council, Industrial Health Research Board, London.
- DE SILVA, H. R. On an investigation of driving skill. *Hum. Factor, London*, 1936, 10, 1-13.
- GRIGG, A. E., & THORPE, J. S. Deviant responses in college adjustment clients: A test of Berg's deviation hypothesis. *J. consult. Psychol.*, 1960, 24, 92-94.
- HARRIS, J. L. Deviant response frequency in relation to severity of schizophrenia. Unpublished master's thesis, Louisiana State University, 1958.
- HESTERLY, S. O., & BERG, I. A. Deviant responses as indicators of immaturity and schizophrenia. *J. consult. Psychol.*, 1958, 22, 389-393.
- JONES, H. E., & CONRAD, H. S. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monogr.*, 1933, 13, 223-298.
- KATZELL, R. A. Cross-validation of item analyses. *Educ. psychol. Measmt.*, 1951, 11, 16-22.
- MAINLAND, D., & MURRY, I. M. Tables for use in fourfold contingency tests. *Science*, 1952, 116, 591-594.
- MILES, W. R. Measures of certain human abilities throughout the life span. *Proc. Nat. Acad. Sci. Wash.*, 1931, 17, 627-633.
- MILES, W. R. Abilities of older men. *Personnel J.*, 1933, 11, 352-357.
- PEARSON, K. On our present knowledge of the relationship of mind and body. *Ann. Eugen. Cambridge*, 1925-1926, 1, 382-391.
- ROITZSCH, J. C., & BERG, I. A. Deviant responses as indicators of immaturity and neuroticism. *J. clin. Psychol.*, 1959, 15, 414-417.
- RUCH, F. G. The differentiative effects of age upon human learning. *J. gen. Psychol.*, 1934, 11, 261-286.
- RUGER, H. A., & STOEISSIGER, B. Growth curves of certain characteristics in man (males). *Ann. Eugen., Cambridge*, 1927, 2, 76-110.

(Received July 5, 1962)

# CONSTRUCTION AND VALIDATION OF THE FOOD ATTITUDE SCALE<sup>1</sup>

DONN BYRNE, CAROLE GOLIGHTLY, AND E. J. CAPALDI<sup>2</sup>

*University of Texas*

A scale designed to measure a personality dimension involving attitudes toward food was constructed by means of internal-consistency item-analysis procedures. A pool of 221 true-false items was administered to 400 undergraduates. The analysis yielded 62 cross-validated items scored for males, 65 for females. With new samples, split-half reliability was found to be .73 for males, .74 for females. Test-retest reliability (6 weeks) was .86 for males, .82 for females. In a validation study with 204 undergraduates, scores on the food scale were found to be positively correlated with self-ratings of viscerotonia and negatively correlated with cerebrotonia, as hypothesized.

Investigations of individual differences in affective responses to food or food-related cues have been confined primarily to aversive reactions. Typically, a list of various foods is presented, and subjects are asked to indicate those items which they dislike, those which they would refuse to eat, etc. With aversion to food conceptualized as a personality variable, it has been found that the number of foods disliked is positively related to a number of relatively negative characteristics: feeding problems (McCarthy, 1935), neuroticism (Gough, 1946; Smith, Powell, & Ross, 1955a; Wallen, 1945), various other indices of maladjustment (Altus, 1949; Davids & Lawton, 1961; Wallen, 1948), immaturity and a general tendency to indicate dislikes (Lindgren, 1962).

Even though psychological research dealing with food attitudes has largely been limited to food aversions, it seems obvious that a considerably broader array of attitudes concerning food and eating should constitute

a personality dimension of possible interest and usefulness. Wallen (1945) suggests that food habits reflect many family practices and attitudes. It would be expected, then, that the recurring events associated with eating throughout infancy and childhood are related not only to adult attitudes toward food, but also to a number of other attitudes. For example, there is some evidence that the greater the variety of foods which children like, the more positive is their self-concept and their mother-concept (Davids & Lawton, 1961). In addition, early food experiences have long been discussed by personality theorists as important determiners of a series of characteristics related to "oral dependency" and similar formulations.

In view of these considerations, a scale which measures general food attitudes would seem to be a potentially useful tool in personality research. The construction of such a scale and the determination of its reliability are described in the following section, and an initial validation study is reported in a later section.

## CONSTRUCTION OF THE SCALE

### *Initial Item Selection*

The intent was to construct a series of True-False items which would cover an array of eating experiences. A series of 221 items were written, and these were grouped into three categories. The first, Past Attitudes and Habits, contained items dealing primarily with childhood events. Examples are:

15. Discipline was usually enforced shortly before or after the evening meal.

17. My father enjoyed eating.

<sup>1</sup> The Food Attitude Scale, Behavior Rating Scale, and scoring keys for each have been deposited with the American Documentation Institute. Order Documentation No. 7506 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$2.50 for microfilm or \$1.75 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> The authors wish to express their appreciation to James Barry, John Brelsford, Fred Labowitz, and Carl McGraw, Gladene Michel, John Sheffield, and Jerry Simmons who served as research assistants on this project.



25. Less than an average amount of conversation occurred at mealtime in my family.

28. Business matters were often discussed at meals (chores, etc.).

31. Sometimes my mother would give me my favorite food when I was sick or unhappy.

The second section, Food Preferences, more or less followed the form of the various food aversion tests. Examples are:

65. I like apricots.

92. I like chili.

110. I like green salad.

120. I like liver.

145. I like sardines.

The third section, Present Attitudes and Habits, dealt with current behavior. Examples are:

171. In general, I prefer a slow leisurely meal to a quick, hurried one.

177. I do not care much for desserts.

186. A good wife must be a good cook.

206. Watching people eat makes me hungry.

208. I often buy refreshments at movies, ball games, etc.

The items were arranged in a booklet entitled *Survey of Eating Habits*; responses were recorded on IBM answer sheets.

### Item Analysis

Because there is no obviously acceptable external criterion of food attitudes, an internal consistency item analysis was employed. A number of investigators have reported significant sex differences in the food aversion studies (Smith et al., 1955a; Smith et al., 1955b; Wallen, 1943), and it was decided to perform separate item analyses for male and female subjects.

An initial scoring system was devised in the following manner. Two of the authors (DB and EJC) independently designated which of the 221 items appeared to be the most definitive measures of a dimension of food attitudes. There was agreement on 50 such items, and these constituted the initial scoring key for the test.

The 221-item questionnaire was administered to 400 students (200 males, 200 females) enrolled in the introductory psychology course at the University of Texas. Sample I consisted of 100 males and 100 females. For each sex, a frequency distribution was plotted, and those subjects falling in the upper and lower 27% of the distribution were the criterion groups in each item analysis (Kelley, 1939). Utilizing Flanagan's Table (Thorndike, 1949, pp. 348-351), product-moment correlations between responses to each item and total score were determined. This procedure was repeated for Sample II (100 of each sex) as a cross-validation.

For each sex, those items which yielded correlations significant beyond the 5% level in both samples were designated as scorable items in the final scale. For males, 62 of the 221 items met this criterion. For females, 26 of the same items reached statistical significance in both samples along with

an additional 39 items. Thus, the final scoring system is based on 101 items.

### Reliability

A new sample of 73 male and 86 female students was drawn. The 221-item Food Attitude Scale (FAS) was administered, and coefficients of internal consistency were computed for each sex. For males scores on odd and even items correlated .73, corrected by the Brown-Spearman formula. For females, the corrected split-half reliability coefficient was .74.

A portion of this same sample (41 males, 55 females) took the FAS a second time, after a period of 6 weeks. Test-retest comparisons yielded coefficients of .86 for males and .82 for females. Thus, the instrument is sufficiently reliable for research use; however, the magnitude of these reliability coefficients indicates that further work on item selection is warranted in order to increase consistency of measurement.

### Discussion

Any meaningful discussion of the personality dimension measured by this instrument must await the accumulation of correlational and experimental data. However, it may be of value to examine briefly the content of the items which comprise the final scale. Fully realizing the dangers involved in naively assuming that item content reflects even the phenomenological world of subjects, the following generalizations are presented tentatively as a summary of the material contained in the scored items and as a possible source of hypotheses.

Despite differences in specific content, males and females are quite similar in the kinds of items which constitute this attitudinal dimension. Individuals who are on the positive (food liking) end of the continuum, in contrast to those on the negative (food disliking) end, tend to indicate that mealtime during their childhood was a pleasant social experience in which all members of the family participated in an unhurried and often elaborate (flowers or candles on the table) ceremony. Light conversation took place during and after eating, without unpleasantness such as arguments, disciplining, scolding, or assignment of chores. All members of the family liked to eat, and the mother was perceived as a good cook who enjoyed this particular household task. In addition, food seemed to be utilized by parents as a reward in that favorite foods

were prepared when the child was sick or unhappy or on special occasions such as birthdays; going to a restaurant was a special treat.

With respect to present attitudes, those on the positive end of the dimension still regard mealtime as a pleasant activity, and food is apparently utilized by these individuals as a self-administered reward (celebrating special events by going to a restaurant, eating when depressed, etc.). Many kinds of activities are made to include eating in that these individuals say that they buy refreshments at movies and ball games, eat peanuts at parties, like to smell food cooking, eat while watching TV, and both males and females agree that a good wife must be a good cook. They also indicate that they like desserts, sometimes crave sweets, like foreign dishes, and like to try new foods. In addition, they say that they tend to gain weight, are talkative, that children should not be forced to eat, and that indigestion and heartburn are rare. Thus, favorable present attitudes toward food-oriented activities appear to be related to a history of positive reinforcement associated with food and eating.

Specific food preferences are difficult to interpret meaningfully. In this section of the test, items which differentiated those high and low in food liking exhibited the least amount of overlap for male and female subjects. Only three of the food preference items differentiated high and low scorers of both sexes: chicken, hot dogs, and lettuce.

With respect to sex differences per se, other investigators have reported that females have significantly more food aversions than males (Smith et al., 1955a, 1955b). Wallen (1943) found that among subjects of low socioeconomic status females disliked more foods than males, but he found no sex differences for subjects of high socioeconomic status. On the present test for a sample of 106 females, the mean number of food aversions (out of 104 foods) was 21.02; for a sample of 98 males the mean was 19.01. This difference was not statistically significant ( $t = .16$ ,  $df = 202$ ). Another sample of 50 males and 50 females was selected for a sex-difference item analysis of the food list. As shown in Table 1, 15 of the foods yielded

TABLE 1  
ITEMS YIELDING SIGNIFICANT SEX DIFFERENCES  
IN FOOD AVERSION

Food	Percent- age of males indicating dislike	Percent- age of females indicating dislike	$\chi^2$
Calves' brains	70	90	6.25**
Celery	32	8	9.00***
Clam dip	46	74	8.17***
Eggs	4	18	5.01*
Hominy	28	52	6.00**
Kidney stew	80	94	4.33*
Lamb	20	38	3.93*
Left-overs	24	44	4.46*
Nuts	2	14	4.89*
Oysters	26	64	14.59****
Tripe	64	96	16.00****
Turtle soup	60	88	10.19***
Waffles	4	18	5.01*
Watermelon	2	12	3.84*
Whole wheat bread	6	22	5.32*

\* $p < .05$ .

\*\* $p < .02$ .

\*\*\* $p < .01$ .

\*\*\*\* $p < .001$ .

significant sex differences. In only one instance (celery) did males indicate greater dislike than females.

The relatively low reliability of the FAS suggests that the test is not a homogeneous measuring instrument. It was found that when the three sections of the test are scored separately, the highest relationships for each sex were between Past Attitudes and Habits and Present Attitudes and Habits. As shown in Table 2, Food Preference scores were unrelated to either of the other sections in a group of 104 females and related only to Present Attitudes and Habits in a group of 98 males.

#### TEMPERAMENT TYPES AND ATTITUDE TOWARD FOOD

##### Background

Many personality theorists have described personality traits or types in which food-related behavior, or more generally orality, plays a prominent part. For example, Freud and Abraham describe the oral-passive type, and there is Fromm's receptive orientation, Adler's pampered child, and Horney's com-



TABLE 2  
INTERCORRELATIONS OF THE THREE SECTIONS  
OF THE FOOD ATTITUDE SCALE

	Males		Females	
	Past attitudes and habits	Present attitudes and habits	Past attitudes and habits	Present attitudes and habits
Past attitudes and habits	.10	.35***	-.16	.23*
Food preferences		.32***		-.06

\* $p < .05$ .  
\*\*\* $p < .01$ .

pliant character. The behavior proposed by these various analysts for such individuals includes a tendency to gain support from parents, friends, authorities, and God; a tendency to be optimistic, friendly, and loving when things go well and anxious and distraught when rebuffed; the seeking of consolation in eating and drinking because being fed is equated with being loved; conforming behavior; and a desire to please and succeed without effort (Munroe, 1955). These clinically derived observations suggest a series of studies relating the FAS to these various oral characteristics.

A first step, however, would seem to be that of determining whether or not scores on this scale are, in fact, related to a more general personality variable of orality. One personality theorist who has described an oral dimension is Sheldon (1942). It is possible to utilize Sheldon's definitions of the "three primary components" of temperament for research purposes without any assumptions about possible interrelationships between somatotype and temperament. Obviously, his viscerotonic characteristic constitutes the sort of personality dimension with which the food scale would be expected to correlate. He has defined 20 rather specific traits which make up each of his temperament components, thus facilitating the construction of an appropriate measuring device.

The viscerotonic component is characterized by relaxation, love of comfort, an oral

orientation, sociability, affectionate behavior, etc. If these characteristics do form a general personality trait, the more specific dimension of food attitudes should be related to it. Therefore, it is hypothesized that the FAS is positively related to viscerotonia.

The somatotonic component is characterized by assertiveness, energy, a muscular orientation, dominance, courage, aggressiveness, competitive behavior, etc. There does not appear to be any a priori reason for predicting either a positive or a negative relationship between these traits and food attitudes.

The cerebrotonic component is characterized by restraint, overintensity, a cerebral orientation, need for privacy, sensitivity, introverted behavior, etc. With respect to food, Sheldon (1942) suggests that for viscerotonics, the food-taking time is the high spot of the day while cerebrotonics want to eat quickly and have it over with; cerebrotonics care little for service and ceremony or for fancy dishes. It is hypothesized that the FAS is negatively related to cerebrotonia.

### Constructing a Temperament Scale

Attempts to measure Sheldon's temperament variables have been of two major varieties. His original procedure involved ratings on a 7-point scale made by an examiner who observed the subjects closely for at least a year in many different situations and conducted not less than 20 analytic interviews.

For most research purposes, such an approach is impossibly forbidding. In a subsequent investigation, Child (1950) constructed a 66-item multiple-choice questionnaire to measure these temperament components.

For the present validation study, two of the authors (CG and DB) constructed a 60-item self-rating scale which was designed to follow Sheldon's trait descriptions as closely as possible. Examples are:

#### Viscerotonia

1. My responses are slow and deliberate (both verbal and motor reactions).

28. I like to be surrounded and supported by others (I love company, enjoy being one among many, feel a genuine fondness for people in general).

37. I freely communicate my feelings (seldom hold things back, little emotional inhibition, little repression of emotion).

47. I think childhood is perhaps the best period of life (more fun to be a child than an adult, I enjoy children and get along with them).



TABLE 3

SPLIT-HALF RELIABILITY COEFFICIENTS  
FOR THE TEMPERAMENT SCORES

	Males	Females
Viscerotonia	.55	.68
Somatotonia	.76	.80
Cerebrotonia	.20	.56

#### Somatotonia

10. I seek and enjoy physical adventure for its own sake (like speed, danger, horses, large dogs, rough sports).

18. I love being out in the open (dislike small rooms and places in which it is difficult to move freely, like large rooms, life on a grand scale, houses on hills).

33. Under the influence of alcohol I become more openly and uninhibitedly aggressive, more expansive, and am full of a sense of power.

50. I regard the period of early manhood or womanhood as the best time of life.

#### Cerebrotonia

6. I feel that the greatest happiness lies in the later decades of life when an individual is relatively free from tenseness and emotional insecurity.

19. I prefer to be alone when faced by trouble or grief.

32. I am physiologically overresponsive (easily nauseated, poor digestion, frequently constipated).

39. I am very aware of everything around me and often have a feeling of apprehensiveness.

Items representing the three components were arranged in random order. For each item, subjects responded on a 5-point frequency scale: Almost Never, Infrequently, Sometimes, Frequently, Usually.

### Procedure

Both the Food Attitude Scale and the temperament scale (labeled "Behavior Rating Scale") were administered on separate occasions to a sample of 204 students (98 males, 106 females) enrolled in the introductory psychology course at the University of Texas.

### Results

First, the reliability of the three temperament scores was investigated. As may be seen in Table 3, the corrected odd-even correlation coefficients reveal that the internal consistency of these temperament scales is relatively low; only for somatotonia do they approach adequacy.

Despite the magnitude of these reliability coefficients, scores on the FAS were correlated with each temperament variable, separately for males and females. In Table 4 these coefficients are given. As hypothesized, scores on the FAS are positively related to viscerotonia and negatively related to cerebrotonia. Somatotonia and FAS were unrelated in males, as predicted, but an unexpected positive relationship was found for females.

In addition, separate correlations with the temperament variables were obtained for each of the three sections of the FAS as is also shown in Table 4. With only one exception (females for the cerebrotonia component), the relationships are accounted for by the Past Attitudes and Habits and by

TABLE 4  
RELATIONSHIPS BETWEEN FOOD ATTITUDES AND TEMPERAMENT COMPONENTS

	Males			Females		
	Viscero-tonia	Somato-tonia	Cerebro-tonia	Viscero-tonia	Somato-tonia	Cerebro-tonia
Food attitude scale	.21*	.09	-.29***	.36***	.27***	-.22*
Past attitudes and habits	.22*	.08	-.26***	.32***	.22*	-.18
Food preferences	.02	.03	-.16	.11	-.06	-.26***
Present attitudes and habits	.22*	.08	-.20*	.23*	.20*	-.04

\*  $p < .05$ .  
\*\*\*  $p < .01$ .

TABLE 5

CORRELATIONS BETWEEN TEMPERAMENT SCALE ITEMS AND FOOD ATTITUDE SCALE SCORES

<i>r</i> for males	<i>r</i> for females	Item
.34***	.41***	15. I take pleasure in eating and the ceremony connected with eating (well-set table, company, candles).
-.27***	-.32***	30. I do not sleep at night and feel fatigued during the day (difficulty in going to sleep, light sleeper, wake easily in the morning).
.25*	.35***	45. My sleep is deep, easy, and undisturbed (go to sleep quickly, sleep through the night).
-.22*		5. I have a strong need for privacy (I would like to be cut off from all but one or two intimate associates).
-.21*		16. I prefer to remain silent about my feelings even when they are intense (dislike outward expression of joy, sorrow, etc.).
.21*		26. I enjoy competition (put myself forward, not afraid to ask questions, not easily embarrassed).
-.21*		36. I dislike being socially involved (avoid social gatherings, dislike comradeship beyond a few intimates).
.22*		41. I like to do what is accepted and approved (prescribed routine, polite manner of approach to what I seek, the ceremony related to such events as births, marriages, and holidays is important to me).
.20*		48. I feel the need for people when troubled (seek people in the face of stress, feel better after telling others a problem, seek companionship during times of sorrow).
.23*		57. I do things heartily, thereby making noise (in walking, moving, speaking, laughing, coughing).
	.21*	2. I am relatively complacent (see little hurry or urgency in any situation, cross bridges when I come to them, let sleeping dogs lie).
	.27***	9. I turn to action when confronted with trouble or tragedy.
	.21*	11. I bear injury and pain lightly (relatively oblivious to physical strain, find cold showers stimulating, do not flinch when I get a shot).
	.23*	17. I have an intent and sensitive-appearing face (all activity of facial musculature is rapid and alert).
-.20*		21. People say that I look younger than I am.
	.20*	24. I am relaxed and without tension.
	.30***	25. My voice is unrestrained and carries well.
	.26***	27. I have few mood changes (feelings are steady, dependable, and unchanging, harmony and evenness of outlook).
-.20*		35. I am sensitive to pain (dislike doctors and dentists intensely, dislike cold showers, apprehensive of anticipated pain).
	.21*	38. I feel alert and active (sit, stand, and walk erectly and vigorously without effort, use vigorous gestures).
-.20*		49. I am not very intense in mental and moral outlook (lack of strong drive except for the elementary biological needs, do not get excited about most issues).
	.22*	54. I lack uniformity in behavior and outlook (attitudes are subject to a sudden and disconcerting change, abrupt changes of mind).
-.20*		58. I am not particularly sensitive to the less obvious or more subtle needs of others.
	.24*	60. People say that I look older than I am.

\*  $p < .05$ .\*\*\*  $p < .01$ .

the Present Attitudes and Habits rather than the Food Preferences section.

One further analysis was undertaken. Each of the 60 items of the temperament scale was correlated with the FAS for males and females separately. The items yielding the highest coefficients are shown in Table 5. Once again sex differences are apparent, suggesting that this oral dimension has a somewhat different meaning in males and females. It would appear that male subjects fit the description of an "oral character" somewhat more closely than females.

### *Discussion*

The hypotheses were confirmed, but the magnitude of the correlations is disappointingly low. Even discounting the low reliability of these research instruments and correcting for attenuation, the resulting coefficients are still relatively small. It is possible that the various elements proposed as components of an oral personality dimension are less highly related than various theoretical formulations would suggest. It may be that food attitudes are highly related to a few "oral characteristics" and unrelated to others.

### IMPLICATIONS FOR FURTHER RESEARCH

Additional work in refining the Food Attitude Scale would be desirable. While it is sufficiently reliable for exploratory research, greater consistency of measurement should be sought in further refinements of the scale. The low interrelationships of the three sections of the scale suggest that this dimension is not a unitary one; a factor analysis should yield more homogeneous dimensions of attitudes toward food.

Other investigations with this personality dimension are suggested by previous research in this area. First, there has been a long-standing interest in child-rearing practices in relation to eating problems. Several items on the FAS suggest that the families of food likers have a permissive, rewarding, and relaxed family orientation toward mealtime activity. And studies of parental behavior such as that of Sears, Maccoby, and Levin (1957) indicate that the mothers of children

with feeding problems tend to be relatively punitive, cold, and rejecting. Thus, the study of the antecedents of food attitudes in parent-child interactions would seem to hold promise.

A second area of potential interest is that of the oral character and the personality variables which various theorists suggest should be interrelated. In the present work, modest relationships were found between FAS scores and viscerotonia and cerebrotonia. Other aspects of orality may also be found related to food attitudes. For example, Blum and Miller (1952) measured orality in children by counting nonpurposive mouth movements in a series of time sampling procedures. They found that orality was positively related to consumption of ice cream over a 3-week period, ratings by teachers of eagerness at lunch time, ratings of eagerness to have other children like them, dependency on adults and other children, and taste suggestibility, and negatively related to popularity. It is hypothesized, then, that food attitudes should be related to the sorts of behaviors investigated by Blum and Miller as well as to affiliation need, religious beliefs, alcohol consumption, conformity, predisposition to ulcer development, etc. Because of the consistency with which sex differences were found in the present research, it may be well to investigate these relationships with orality separately for males and females.

Not only are those high on the oral dimension supposed simply to enjoy food and drink, but it has often been hypothesized that they tend to respond to anxiety and depression by eating and drinking (Fromm, 1947). Again, several items on the FAS lend support to this general proposition. Thus, the psychological components of obesity may combine strongly positive food attitudes and anxiety. Conversely, extreme underweight could involve strongly negative food attitudes plus anxiety. In a situation in which anxiety was aroused, differential predictions would be made about the food intake of those scoring at opposite extremes of the FAS.

Any situation in which hunger is aroused would conceivably lead to different response patterns on the part of those at opposite



extremes on the food attitude dimension. For example, under deprivation conditions food likers, in contrast to food dislikers, might be expected to produce more food-related imagery in thematic apperception productions (Atkinson & McClelland, 1948), to be more sensitive to food stimuli presented subliminally (Byrne, 1959), to learn and to recall food-related words more readily (Epstein & Levitt, 1962), and to have greater difficulty in adhering to restrictive diets (Summerskill & Darling, 1955).

## REFERENCES

- ALTUS, W. D. Adjustment and food aversions among Army illiterates. *J. consult. Psychol.*, 1949, 13, 429-432.
- ATKINSON, J. W., & MCCLELLAND, D. C. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, 38, 643-658.
- BLUM, G. S., & MILLER, D. R. Exploring the psychoanalytic theory of the oral character. *J. Pers.*, 1952, 20, 287-304.
- BYRNE, D. The effect of a subliminal food stimulus on verbal responses. *J. appl. Psychol.*, 1959, 43, 249-252.
- CHILD, I. L. The relation of somatotype to self-ratings on Sheldon's temperamental traits. *J. Pers.*, 1950, 18, 440-453.
- DAVIDS, A., & LAWTON, MARCIA J. Self-concept, mother-concept, and food aversions in emotionally disturbed and normal children. *J. abnorm. soc. Psychol.*, 1961, 62, 309-314.
- EPSTEIN, S., & LEVITT, H. The influence of hunger on the learning and recall of food related words. *J. abnorm. soc. Psychol.*, 1962, 64, 130-135.
- FROMM, E. *Man for himself*. New York: Rinehart, 1947.
- GOUGH, H. G. An additional study of food aversions. *J. abnorm. soc. Psychol.*, 1946, 41, 86-88.
- KELLEY, T. L. The selection of upper and lower groups for the validation of test items. *J. educ. Psychol.*, 1939, 30, 17-24.
- LINDGREN, H. C. Age as a variable in aversion toward food and occupation. *J. consult. Psychol.*, 1962, 26, 101-102.
- MCCARTHY, D. Children's feeding problems in relation to the food aversions in the family. *Child Developm.*, 1935, 6, 277-284.
- MUNROE, RUTH L. *Schools of psychoanalytic thought: An exposition, critique, and attempt at integration*. New York: Dryden Press, 1955.
- SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.
- SHELDON, W. H. *The varieties of temperament*. New York: Harper, 1942.
- SMITH, W. I., POWELL, ELIZABETH K., & ROSS, S. Food aversions: Some additional personality correlates. *J. consult. Psychol.*, 1955, 19, 145-149. (a)
- SMITH, W. I., POWELL, ELIZABETH K., & ROSS, S. Manifest anxiety and food aversions. *J. abnorm. soc. Psychol.*, 1955, 50, 101-104. (b)
- SUMMERSKILL, J., & DARLING, C. D. Emotional adjustment and dieting performance. *J. consult. Psychol.*, 1955, 19, 151-153.
- THORNDIKE, R. L. *Personnel selection*. New York: Wiley, 1949.
- WALLEN, R. Sex differences in food aversions. *J. appl. Psychol.*, 1943, 27, 288-298.
- WALLEN, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, 40, 77-81.
- WALLEN, R. Food aversions in behavior disorders. *J. consult. Psychol.*, 1948, 12, 310-312.

(Received July 25, 1962)

## VERBAL CONDITIONING OF SCHIZOPHRENICS AND NORMALS IN A SITUATION RESEMBLING PSYCHOTHERAPY

JOAN SLECHTA, WILLIAM GWYNN, AND CROCKER PEOPLES<sup>1</sup>

*Fort Hays Kansas State College*

10 normal Ss and matched Ss diagnosed "schizophrenic" were asked individually to tell two stories including as characters themselves, the experimenter, and two other people. During the first story E was as neutral as possible, but in the second, E reinforced references to the experimenter (the criterion response) using both verbal and gestural cues ("mmm-hm," nods of the head, etc.). The null hypothesis was tested and rejected. Normals not only emitted a significantly greater percentage of criterion responses under reinforcement than did the hospitalized Ss, but also conditioned more thoroughly ( $p = .01$ ). Implications for psychotherapy were discussed.

The present study was undertaken to investigate differences existing between hospitalized (NP) subjects and normal subjects with regard to a type of verbal conditioning in a situation resembling psychotherapy. Elsewhere in the literature, one can find numerous studies of verbal conditioning, many employing rather ingenious devices to measure the "conditionability" of various groups of subjects (see, for example, the review of Salzinger, 1959). Few studies, however, can be found which satisfy both goals of the present undertaking: (a) to compare hospitalized and normal subjects on a verbal conditioning task; and (b) to relate this to a psychotherapeutic type of situation.

Of the numerous reports on verbal conditioning in which the familiar "mmm-hm" or "yes," etc. are employed, those which appear to be most readily identified with a psychotherapeutic situation are the investigations which have utilized conversation, storytelling, and/or interviews. Verplanck (1955), for example, found it to be possible to control the content of a conversation through reinforcement. McNair (1957) in-

creased the rate of verbalization while subjects were talking about photographic slides. Mock (1957) had schizophrenic subjects tell 20 10-minute stories and found in general an increase in positively reinforced responses and a decrease in negatively reinforced responses. Krasner (1958), employing a similar method, found that the schizophrenic subjects' references to a preselected class of words increased or decreased in relation to the experimenter's reinforcement. Salzinger and Pisoni (1958) showed that it was possible to condition self-referred affect statements in both normal and schizophrenic subjects during an otherwise standard clinical interview.

Several writers (Dollard & Miller, 1950; Pascal, 1959; Wolpe, 1958) have reported behavioral changes during psychotherapy through the application of such learning principles as counter conditioning, generalization, discrimination, extinction, and various methods of reward or reinforcement. The problem which seems to exist for most learning-oriented psychotherapists is what to reinforce and how to reinforce it. That most all psychotherapists employ some of the above principles in therapy, perhaps unwittingly, is obvious. Dinoff, Rickard, Salzberg, and Sipprelle (1960) pointed out that in a psychotherapeutic situation, a person's remarks could be classified into three general groups: references to his environment, references to himself, and references to the thera-

<sup>1</sup> This report is based on a study submitted by the first author as a master's thesis to the Graduate Faculty of the Fort Hays Kansas State College, May, 1962.

The writers would like to thank Steve Pratt, Chief Psychologist, and others at Larned State Hospital for their aid and cooperation in this study.

TABLE 1  
SUBJECT VARIABLES

Experimental subjects					Control subjects				
Subject	Age	Sex	Marital status	Years in school	Subject	Age	Sex	Marital status	Years in school
1	29	F	M	14	1	29	F	M	14
2	29	F	M	14½	2	29	F	M	14½
3	27	F	D	12	3	26	F	M	12
4	27	F	D	12	4	27	F	M	12
5	27	F	M	12	5	27	F	M	12
6	27	F	M	12	6	27	F	M	12
7	27	F	M	12	7	27	F	D	12
8	19	M	S	12	8	19	M	S	12
9	28	M	S	14½	9	28	M	S	14½
10	30	M	S	15	10	30	M	S	15

pist. Various types of therapy emphasize the importance of one or another of these areas, but regardless of the focus, any psychotherapeutic relationship is dependent upon the client's ability to respond to the therapist and the extent to which he can use the therapist's reactions to modify his own behavior.

Dinoff et al. hypothesized that the success a therapist has in establishing a relationship with a patient might be determined by the ease with which the patient progresses from references to the environment to himself to the therapist. Enlarging on this hypothesis, it is possible that the extent to which an individual is capable of entering productively into therapy might be determined by his ability to be conditioned to include the therapist in his verbalizations. This might be characterized by a rise in the frequency of references to the therapist.

In an attempt to duplicate the conditions of a psychotherapeutic situation, a group of subjects diagnosed "schizophrenic" and a group of matched normals were seen individually by the same experimenter and asked to tell two stories. An effort was made to increase the frequency of occurrence of one class of verbal behavior in the second story. The null hypothesis of no difference between groups with regard to conditioning was tested.

## PROCEDURE

*Subjects.* The subjects for this study were three male and seven female patients at Larned State Hospital, Larned, Kansas, who had been diagnosed as chronic schizophrenic and three male and seven female normal control subjects. The control subjects were matched as closely as possible to the schizophrenics with regard to age, sex, and education. Table 1 gives a short resumé of the subject variables. All had a high school education or better and the median age of both groups was 27.

*Method.* The method used is similar to that which Dinoff et al. (1960) employed with normal subjects. Before any formal instructions were given to the subject, the experimenter entered into conversation with him. Since lengthy instructions might have been confusing to the experimental subjects, some instructions were included in the course of this conversation to both groups. In this conversation, each subject was told that he would be asked to tell two stories employing any plot or setting desired and that the stories would be tape recorded. The following instructions were then read to the subject:

I would like to have you tell two stories including you, me, and two other people as characters. Make them about 5 minutes long. Any questions? Remember to include you, me, and two other people in the story.

While the subject was telling the first story, the experimenter remained silent and motionless. During the second story the subject was reinforced by the experimenter each time he included the experimenter in his story. Reinforcement consisted of head nodding, leaning forward toward the subject, smiling, and/or remarks such as "mmm-hm," "good," etc. These were administered in a variable order in an effort to render the situation more realistic, i.e.,



more like a psychotherapy session. In case a subject ended either story short of 2 minutes, the experimenter said, "Please make the story longer," or "Tell me more."

Each story was presented in typed form, with phrases numbered and divided by red slashes, to two judges who were naive regarding the purpose or rationale of the study. The judges (two students) were instructed in terms of concise, unambiguous criteria to consider each separate phrase in each story and determine to which of the following categories it pertained: (a) Environment—E; (b) Experimenter—T (for therapist); (c) Subject—S; or (d) Ambiguous—A. Scoring sheets were provided for each story and the judges independently designated each phrase as E, T, P, or A. The stories were coded so the judges would be unable to ascertain which group they were from and were presented to the judges with the following instructions which follow closely those developed by Dinoff et al. (1960):

The stories you are to score should include at least four characters. They are: the person telling the story, whom we shall designate as (P); the examiner or person listening, whom we shall designate as (T); and at least two other people whom we shall designate as (E). In addition, all people other than (P) and (T) and all things in the environment and situations other than those including (P) or (T) shall be designated as (E).

You are to score these stories, statement by statement, to determine which numbered statements are about (P), (T), or (E). The following four scoring categories are to be used as the score which you will record on the scoring sheet which will be provided:

1. (T) When a statement includes any reference to the examiner (or listener) including him alone (T), or him in connection with the story teller (P, T), or (T, E)—score "T", i.e., any reference to the listener.
  2. (P) When a statement includes any reference to the (P) alone or to (P, E)—score "P", i.e., any phrase excluding the listener.
  3. (E) when a statement does not include (P) or (T)—score "E", i.e., others in the story.
  4. When a statement is unclear as to just who it includes it is ambiguous and is to be scored "A".
- Remember to score (T) if there is any reference to the listener and this includes any phrase or statement in which a personal pronoun referring to the listener is understood from the context of the phrase.

## RESULTS AND DISCUSSION

Of the total of 2,039 responses given by all subjects, two judges agreed on 2,033, or 98.23%, indicating a highly reliable scoring technique. For the remaining 36 responses a third judge was asked to perform classifications and the judgment of the two judges

who agreed on these 36 responses was used in computation.

The data were analyzed in terms of frequency of occurrence of the criterion response in each story, with the criterion response being reference to the experimenter. Because the stories varied somewhat in length, the percentage of references to the experimenter in relation to total number of phrases was used in analysis. The binomial expansion and the Wilcoxon *T* test were employed.

Table 2, Columns A and B, shows the number of criterion responses given by each subject at the operant level. It is seen that all normal subjects made at least one reference to the experimenter while three schizophrenic subjects made no such reference. The greater variability of the schizophrenic group is apparent from an inspection of the ranges which were 1–15 criterion responses for normals and 0–24 for schizophrenics. It might be argued that there is justification for disregarding the data of Experimental Subjects 3, 5, and 8 because they made no reference to the experimenter. If this is done and the corresponding control subjects are discarded, there is even less difference between the two groups.

Table 2 compares the frequency and percentage of criterion responses emitted by normals and schizophrenics under conditions of nonreinforcement and reinforcement. It is apparent that schizophrenic subjects, although increasing the percentage of their criterion responses slightly, did not change significantly from nonreinforcement to reinforcement, while normals changed significantly in both direction and magnitude. In other words, normals not only emitted a significantly greater percentage of criterion responses under reinforcement than did schizophrenics, but also in comparison with the hospitalized group, normals conditioned much more thoroughly.

A number of interpretations of the observed difference in verbal conditioning could be advanced. An appropriate one is suggested by Hartman (1955) who mentioned the immunity to social stimulation which is frequently seen in schizophrenics. This is con-

TABLE 2  
FREQUENCY AND PERCENTAGE OF T RESPONSES

Pair	Nonreinforced				Reinforced			
	A. Normals		B. Schizophrenics		C. Normals		D. Schizophrenics	
	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
1	1	2.9	24	82.8	22	44.9	2	3.3
2	2	4.2	10	29.4	47	47.0	23	38.3
3	7	11.7	0	0.0	25	39.7	14	46.7
4	15	39.5	2	4.4	21	35.6	5	11.6
5	1	3.6	0	0.0	45	75.0	0	0.0
6	12	20.0	4	10.5	17	28.9	7	11.7
7	8	20.0	17	73.9	27	41.5	3	8.3
8	2	2.4	0	0.0	24	54.5	0	0.0
9	7	20.6	4	9.3	17	60.7	16	31.4
10	4	14.8	1	5.3	20	54.0	2	12.5
<i>M</i>	5.9	13.0	6.2	10.9	26.5	46.2	7.2	13.5

Note.—Comparison of Columns A and B yields binomial  $P = .17$ ; Wilcoxon  $T = 26.0$  (not significant). A and C yields binomial  $P = .01$ ; Wilcoxon  $T = 1.0$ ,  $p < .01$ . B and D yields binomial  $P = .17$ ; Wilcoxon  $T = 20.5$  (not significant). C and D yields binomial  $P = .01$ ; Wilcoxon  $T = 1.0$ ,  $p < .01$ .

sistent with the traditional view of schizophrenics who are said to withdraw from external realities. Immunity to external stimuli implies that overt behavior of schizophrenics is more determined by internal stimuli than is overt behavior of normals. This suggests that the observed difference in conditionability might be explained in terms of a difference in susceptibility of the two groups to positive external stimuli. If positive stimulation from the outside is not important to schizophrenics, it could hardly be expected to alter their behavior to a significant degree. It has been reported by various writers (Bleke, 1955; Buss & Buss, 1956; Pascal & Swensen, 1952) that schizophrenics may be more responsive to negative reinforcement than to positive reinforcement.

There appear to be direct implications from this study to psychotherapy with schizophrenics. Since most traditional psychotherapy depends heavily on verbal exchange and changes in verbal behavior are generally felt to be the *sine qua non* of gross behavioral change, it is readily apparent that inability to condition verbally would greatly lessen an individual's chances of responding to psychotherapy. Of course it cannot be claimed that no conditioning occurred in the schizophrenic group, since 7 of the 10 sub-

jects did show an increase in percentage of criterion responses under reinforcement; however, the increase was arbitrarily labeled as not of significant magnitude to demonstrate an acceptable level of conditioning. Those schizophrenic subjects who showed behavioral change similar to that of the normal group would probably make better candidates for psychotherapy since it has been demonstrated that their verbal behavior is modifiable. It is conceivable that a standard verbal conditioning technique might be developed for use as a screening device in selecting schizophrenic subjects for psychotherapy.

#### REFERENCES

- BLEKE, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1955, 23, 479-498.
- BUSS, A. H., & BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, 52, 283-287.
- DINOFF, M. H., RICKARD, C., SALZBERG, H., & SIPP-RELLE, C. N. An experimental analogue of three psychotherapeutic approaches. *J. clin. Psychol.*, 1960, 16, 70-73.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- HARTMAN, C. H. Verbal behavior of schizophrenics and normal subjects as a function of types of social reinforcement. Unpublished doctoral dissertation, State University of Iowa, 1955.

- KRASNER, L. A technique of investigating the relationship between the behavior cues of the examiner and the verbal behavior of patients. *J. consult. Psychol.*, 1958, 22, 364-366.
- McNAIR, D. M. Reinforcement of verbal behavior. *J. exp. Psychol.*, 1957, 63, 40-46.
- MOCK, J. F. The influence of verbal and behavioral cues of a listener on the verbal production of the speaker. *Dissert. Abstr.*, 1957, 16, 57-58.
- PASCAL, G. R. *Behavioral change in the clinic: A systematic approach*. New York: Grune & Stratton, 1959.
- PASCAL, GERALD R., & SWENSEN, C. Learning in mentally ill patients under conditions of unusual motivation. *J. Pers.*, 1952, 21, 240-249.
- SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, 61, 65-94.
- SALZINGER, K., & PISONI, STEPHENIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, 57, 84-90.
- VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, 51, 668-676.
- WOLFE, J. *Psychotherapy by reciprocal inhibition*. Stanford Univer. Press, 1958.

(Received August 6, 1962)



## FACTOR SCALES FOR THE CALIFORNIA PSYCHOLOGICAL INVENTORY

ROBERT C. NICHOLS

AND

RICHARD R. SCHNELL

*National Merit Scholarship Corporation*

*Purdue University*

Factor scales were constructed to measure the 2 principle factors isolated in a factor analysis of 18 scales of the CPI. Correlations of these scales with a variety of test and behavioral variables suggests that they are similar to factors of neuroticism and extraversion found in previous studies. The scales were called Person Orientation and Value Orientation. The implications of the findings for test theory were discussed.

The California Psychological Inventory (CPI) consists of 18 scales, most of which have been derived by item analysis against such socially relevant criteria as social status, sex, intelligence, ratings of dominance, etc. Many of the scales are correlated substantially with each other, and have been grouped into four general classes by Gough (1957) partially on the basis of their intercorrelations. A factor analysis of the 18 standard CPI scales (Mitchell & Pierce-Jones, 1960) found four factors which accounted for appreciable portions of the variance of each of the scales, and some scales had little reliable variance other than that accounted for by the factors.

The finding of a few factors to account for a substantial portion of the reliable variance among the CPI scales is of particular interest, since most of these scales were constructed by criterion keying. Since each item, by virtue of being included in a scale, has been shown to be related to some important non-test-behavior, it is likely that the factors will also be related to important criterion behaviors. This is not guaranteed, however, for there are three possible sources of correlation between two criterion keyed scales and thus of the factors derived from these correlations: (a) Common variance between scales may represent common variance in the traits the scales purport to measure. One would certainly expect the "folk concepts" which the CPI scales attempt to measure to be correlated in most populations, and this correlation among psychological traits may be responsible for the intercorrelation among the scales. To the extent that this is the case, the

CPI factors would be identical with the trait factors. (b) Common variance between scales may be due to some aspect of item response that is related to the traits which both scales purport to measure, but which is not dependent on any correlation among the measured traits. It is possible that such response characteristics as tendency to describe oneself in favorable terms or tendency to report interpersonal conflicts with others may be related to a variety of otherwise independent behaviors. To the extent that the correlations among CPI scales are due to response tendencies which are common to a variety of independent behaviors, the factors would represent the aspects of item response most likely to be related to any new behavioral criterion. (c) Common variance between scales may be due to common response tendencies which are not related to the traits which the scales purport to measure. It is possible that such response characteristics as tendency to answer true may be unrelated to the behavioral criteria, yet may contribute to error in a number of scales.

Correlations among CPI scales are most likely due to some combination of all three of these common response tendencies. However, to the extent that any one predominates, factor analysis and rotation to simple structure should identify that tendency in its various dimensions as a factor or factors. If it should be the case that there are a few aspects of item response which are related to a great variety of criterion behaviors (the second possibility mentioned above), the factors may be quite useful in predicting new criteria. Whether or not this is the case, a

study of the factors in the CPI and their correlates should contribute to the theory of inventory tests.

The purposes of this study were to factor the CPI, to construct scales to measure the factors, and to obtain preliminary data concerning the correlates of the factor scales.

### PROCEDURE

The correlation matrices of the 18 CPI scales reported in the CPI Manual (Gough, 1957) were factored. Separate factor analyses were done for males ( $N=4,098$ ) and Females ( $N=3,572$ ). These matrices are reported by Gough to be composed of unweighted mean correlations from five separate populations of high school and college students and prison inmates, and in the case of the male matrix, also military officers.

The male and female matrices were factored separately by the principal components method with communalities estimated as the squared multiple correlation of each scale with all other scales. On the basis of the size of the latent roots and the highest loadings for each factor, it was decided to retain three factors from each analysis which were rotated analytically to orthogonal simple structure according to the normalized varimax criterion.

The factors obtained from the male and female matrices were very similar. The first two factors had high loadings of several scales while the third factor had a high loading only of Flexibility, a scale which was not constructed by criterion keying. It was, therefore, decided to obtain factor scales only for the first two factors, since these account for most of the common variance among the criterion keyed scales. In order to obtain factor scores which were relatively independent, high loading scales were selected to represent each factor so that positive and negative loadings on other factors tended to cancel each other. The scales chosen to represent each factor and their loadings are shown in Table 2.

Factor scores for the first two factors were calculated for CPIs from a sample of 300 undergraduate students by summing the standard scores for the representative scales for each factor. These 300 subjects of both sexes had taken the CPI in groups as a part of several previous studies. Factor scales were obtained for both factors using a procedure developed by Webster (1956) for maximizing the correlation of a test with a criterion as follows: The 300 subjects were divided into five groups on the basis of a factor score containing 9%, 20%, 54%, 20%, and 9% of the cases. The number of subjects in each group answering true for each item was tallied, and a  $D$  score was obtained for each of the 480 CPI items by weighting the group total, respectively, +2, +1, 0, -1, and -2. This item analysis was done for each factor, and initial scales for the two factors were obtained by selecting all items which were related to a

given factor at the .01 level and which were not related to the other factor at the .05 level. This was done in order to keep the correlation between factor scales low. This procedure resulted in a 121-item scale for Factor I and a 55-item scale for Factor II. The procedure described by Webster (1956) for discarding items to maximize the correlation of the scales with the criterion was applied which resulted in the discarding of eight items from Factor Scale I and four items from Factor Scale II.

Reliability and factor estimation coefficients of the final scales were determined on several cross-validation groups which were also used to study the relationship of the new scales to other variables. These groups and other data available are as follows:

1. A sample of 57 male and 18 female undergraduate students, who received five or more psychotherapy interviews at the Purdue Psychological Clinic, took the CPI at the termination of therapy. At the same time, their therapist rated them using an Adjective Check List (Gough, 1960).
  2. A sample of 64 high school counselors were tested with the CPI at a counseling training institute.<sup>1</sup> Also available on these subjects were the Strong VIB, the MMPI, the EPPS, the Guilford Zimmerman Temperament Survey, the Minnesota Teacher Attitude Inventory, the Ohio State Psychological Examination, peer ratings on five traits, and  $Q$  sorts from which self-ideal correlations were calculated.
  3. A sample of 250 male college freshmen who were tested with the CPI as part of a research project at their dormitory.<sup>2</sup> Also available on these subjects were the Rotter Incomplete Sentences Blank and dormitory counselors' ratings of adjustment.
  4. A sample of 88 female and 58 male psychiatric patients in an acute intensive treatment hospital were tested with the MMPI which was scored for a 46-item scale for Factor I and a 15-item scale for Factor II. Also available on these subjects were independent interviewer ratings of the patient's symptoms using the Multidimensional Scale for Rating Psychiatric Patients (MSRPP) (Lorr, 1953).
- Relationships between the two CPI factor scales and the other variables listed above were investigated using either point biserial or product-moment correlations.

### RESULTS

The rotated factor loadings for males and females are shown in Table 1. As can be seen from this table, the factor loadings based on male subjects are very similar to those based on female subjects.

<sup>1</sup> The writers are indebted to James Linden for providing these data.

<sup>2</sup> The writers are indebted to John McBrearty for providing these data.



TABLE 1  
ROTATED FACTOR MATRIX

CPI scales	Factor I		Factor II		Factor III		$h^2$	
	M	F	M	F	M	F	M	F
<b>Class I</b>								
Dominance	.28	.23	.63	.74	-.18	-.11	.51	.61
Capacity for status	.37	.39	.60	.68	.20	.21	.54	.66
Sociability	.22	.31	.78	.79	-.14	-.06	.68	.72
Social presence	.00	.10	.71	.73	.27	.36	.67	.58
Self acceptance	-.08	-.06	.74	.77	-.14	.04	.57	.60
Sense of well-being	.70	.78	.23	.25	.07	-.02	.55	.67
<b>Class II</b>								
Responsibility	.68	.67	.10	.18	-.10	-.14	.48	.50
Socialization	.58	.55	-.04	.02	-.25	-.33	.40	.41
Self-control	.83	.87	-.26	-.24	.05	-.15	.76	.84
Tolerance	.71	.74	.20	.32	.34	.22	.66	.70
Good impression	.77	.75	-.05	.05	.06	-.13	.60	.58
Communality	.20	.22	.16	.17	-.38	-.36	.21	.21
<b>Class III</b>								
Achievement via conformance	.79	.75	.19	.23	-.09	-.25	.67	.63
Achievement via independence	.48	.55	.09	.19	.62	.49	.62	.58
Intellectual efficiency	.58	.66	.51	.48	.20	.20	.64	.71
<b>Class IV</b>								
Psychological mindedness	.45	.49	.14	.27	.35	.28	.34	.39
Flexibility	-.01	.01	.02	.05	.71	.67	.50	.45
Femininity	.14	.09	-.24	-.18	-.02	-.27	.08	.11

Table 2 shows the loadings on all three factors of the scales selected as measures of the two factors for which scales were derived. As can be seen, the scales selected to represent a given factor tend to have loadings on the other factors which cancel each other.

Table 3 shows the results of applying Webster's procedure for maximizing the correlation of a test with a criterion. It is obvious from this table that this procedure brings little change in cross-validation reliability, and factorial validity statistics. In this instance, it did not prove to be worth the considerable labor involved in carrying out the analysis.

The data presented in Table 3 show acceptable reliability and factor estimation indices for both scales. It should be pointed out, however, that the factor estimation indices are correlations between the scales and the factor score criteria which have a number of common items and are useful only

as an index of the extent to which the original factor score is approximated by the factor scale.

The two factor scales were relatively independent, correlating .27 in the sample of 250 male freshman students.

#### MEANING OF THE SCALES

There are three lines of evidence which can be pursued for clarification of psychological constructs measured by these two factor scales, namely, the scales or criteria against which the items were selected, the content of the items, and the correlations of the scale score with other test and behavioral variables. Each of the two factor scales will be discussed in terms of these lines of evidence.

#### *Factor I, Value Orientation*

The scales which load highly on this factor are reported by Gough (1957) to measure responsibility, self-control, and maturity, as



TABLE 2  
FACTOR LOADINGS OF SALIENT SCALES USED TO OBTAIN FACTOR SCORES

CPI scales	Factor loadings					
	Factor I		Factor II		Factor III	
	M	F	M	F	M	F
Factor I						
Sense of well-being	.70	.78	.23	.25	.07	-.02
Responsibility	.68	.67	.10	.18	-.10	-.14
Socialization	.58	.55	-.04	.02	-.25	-.33
Self-control	.83	.87	-.26	-.24	.05	-.15
Tolerance	.71	.74	.20	.32	.34	.22
Good impression	.77	.75	-.05	.05	.06	-.13
Achievement via conformance	.79	.75	.19	.23	-.09	-.25
Factor II						
Dominance	.28	.23	.63	.74	-.18	-.11
Sociability	.22	.31	.78	.79	-.14	-.06
Social presence	.00	.10	.71	.73	.27	.36
Self-acceptance	-.08	-.06	.74	.77	-.14	.04

well as a tendency to minimize complaints and a concern over the reactions of others. The items<sup>3</sup> in the scale can be grouped into the following content categories: Emotional stability and control, denial of impulsivity, even temperament and absence of anger and hostility, absence of anxiety and tension, physical health and well-being, dependability and sense of duty, happy home life, respect for others, and respect for rules and social customs. This content and the scale correlates suggest a general psychological ma-

<sup>3</sup>The items and scoring direction for the Value Orientation scale are as follows: 20 F, 22 T, 26 F, 29 F, 30 F, 32 F, 42 F, 44 F, 46 T, 48 F, 49 F, 55 F, 60 F, 71 F, 77 F, 80 F, 81 F, 93 F, 101 F, 103 T, 105 F, 106 F, 109 F, 114 F, 115 F, 117 F, 119 F, 120 F, 127 T, 132 F, 135 T, 138 T, 141 F, 142 F, 149 T, 153 F, 155 F, 157 F, 161 F, 164 F, 165 T, 168 T, 170 F, 174 T, 175 F, 178 F, 181 T, 183 F, 184 F, 185 F, 191 F, 194 F, 195 T, 203 F, 206 F, 209 F, 214 F, 237 F, 247 F, 248 F, 250 F, 253 F, 260 T, 262 F, 266 F, 268 F, 270 F, 275 F, 276 T, 278 T, 279 F, 282 F, 289 F, 291 F, 294 F, 297 F, 298 F, 299 F, 300 F, 302 F, 307 F, 313 T, 336 F, 344 F, 351 F, 353 F, 356 T, 362 T, 367 T, 371 T, 375 F, 380 T, 392 T, 396 F, 398 F, 399 F, 406 F, 407 T, 411 F, 413 T, 419 F, 420 F, 428 F, 431 F, 436 F, 439 T, 451 T, 453 F, 454 F, 463 F, 468 F, 470 F, 474 F.

TABLE 3  
RELIABILITY AND VALIDITY INDICES  
FOR THE TWO FACTOR SCALES

Statistic	Factor I	Factor II
Original scale (before discarding items)		
Number of items	121	55
Reliability (KR-21), original derivation sample	.87	.82
Reliability (KR-21), cross-validation sample <sup>a</sup>	.89	.81
Factor estimation coefficient, cross-validation sample <sup>a</sup>	.96	.88
Final scale		
Number of items discarded by Webster's procedure	8	4
Number of items	113	51
Reliability (KR-21), cross-validation sample <sup>a</sup>	.88	.81
Factor estimation coefficient, cross-validation sample <sup>a</sup>	.96	.89

<sup>a</sup>The correlation between the factor score and the factor scale is here called the "factor estimation coefficient." These were calculated on the cross-validation sample of 250 male college freshmen.

turity reflected in a concern for values and conformity to conventional standards for behavior. It is suggested that this scale be called Value Orientation or V.<sup>4</sup>

The Value Orientation scale correlated positively with a variety of other scales and behavior ratings indicating psychological stability, control, and good interpersonal relations. It correlated negatively with indices of maladjustment and emotionality. Among mental patients, it seemed to be related to an expansiveness—depressive hostility continuum.

With the three exceptions which are noted, all of the following variables were significantly related ( $p < .01$ ) to V. With the MMPI, it correlated positively with *K* (.53) and *L* (.49) and negatively with *Pd* (-.34) and *Sc* (-.38). With the Guilford Zimmerman, it correlated positively with Emotional stability (.44), Objectivity (.70), Friendliness (.58), and Personal Relations (.59). With the Edwards PPS, it correlated positively with Deference (.32) and Endurance (.40) and negatively with Aggression (-.35). It correlated negatively with a maladjustment score on the Rotter ISB (-.31).

Among mental patients, the shorter version of the scale contained in the MMPI<sup>5</sup> correlated positively with psychologist ratings of Expansiveness (.36) and Conceit (.36) and negatively with ratings of Melancholy agitation (-.38), Depressive tension (-.29), Feeling of sinfulness (-.36), Mood swings (-.32), Suicidal preoccupation (-.38), Self-preoccupation (-.31), Lack of faith in self (-.31), Hostile impulses (-.41), and Perception of the world as hostile (-.42).

Among counselor trainees it correlated at the .05 level with peer ratings of Professional attitudes (.31), Improvement of professional

competence (.31), and Extent of peer interaction (.30). High scorers among the clinic clients tended to be rated by their therapists as Conservative, Curious, Dependable, Idealistic, Industrious, Responsible, and Serious, while low scorers were described as Irresponsible, Conceited, and Forgetful.

### *Factor II, Person Orientation*

The scales which load on the second factor are reported by Gough (1957) to measure sociability and favorable interaction with others. The items<sup>6</sup> can be grouped according to content into the following categories: comfortableness with others, joy in interpersonal interaction, dominance and leadership, absence of fear and embarrassment, and quickness of response. This content and the scale correlates suggest that this scale is measuring the familiar Extraversion-Introversion dimension as reflected primarily in interpersonal interaction. It is suggested that this scale be called Person Orientation or P.

The P scale correlated with other scales and behavior ratings which suggested activity and outgoingness. The following are all of the correlations found with this scale which were significant at the .01 level.

In the MMPI, it correlated negatively with *D* (-.36) and *Si* (-.66). With the Guilford Zimmerman, it correlated positively with General activity (.35), Ascendancy (.61), and Sociability (.59) and negatively with Restraint (-.44). With the Edwards PPS, it correlated positively with Exhibition (.37) and Dominance (.39) and negatively with Abasement (-.49). With the Strong, it correlated positively with Personnel Director (.33) and Public Administrator (.26) and negatively with Mathematician (-.35) and Physicist (-.33).

High scores among the clinic clients tended to be described by their therapists as Active,

<sup>4</sup> The scale titles "Value Orientation" and "Person Orientation" were suggested by Harrison Gough.

<sup>5</sup> A shorter version of V can be scored from the 46 scale items which appear in the MMPI. The MMPI booklet numbers and scoring direction for these items are as follows: 3 T, 28 F, 29 F, 36 T, 45 F, 63 T, 68 T, 80 F, 93 F, 96 T, 97 F, 104 F, 117 F, 118 F, 124 F, 139 F, 145 F, 146 F, 157 F, 175 T, 181 F, 198 T, 225 F, 226 F, 251 F, 277 F, 290 F, 306 T, 323 F, 333 F, 348 F, 381 F, 386 F, 395 F, 406 F, 438 F, 459 F, 464 T, 469 F, 472 F, 481 F, 485 F, 516 F, 525 F, 545 F, 558 F.

<sup>6</sup> The items and scoring direction for the Person Orientation scale are as follows: 7 F, 13 F, 25 F, 38 F, 45 T, 52 T, 53 T, 64 F, 68 F, 74 F, 79 F, 85 F, 97 T, 108 T, 131 T, 146 T, 154 T, 163 T, 179 T, 182 F, 186 F, 188 F, 200 T, 202 T, 216 T, 218 T, 223 F, 227 F, 239 T, 242 T, 256 T, 272 F, 287 T, 292 T, 314 F, 319 T, 320 T, 325 F, 346 F, 359 T, 383 F, 385 F, 391 T, 403 T, 412 T, 418 F, 429 F, 448 T, 452 F, 460 F, 475 T.



Energetic, Enthusiastic, and Outgoing, and low scorers were described as Inhibited, Interests narrow, Pessimistic, Dissatisfied, Shy, and Withdrawn.

Among mental patients high scores seem to indicate excitement while low scores indicate feelings of inadequacy. The shorter form of the scale scored on the MMPI<sup>7</sup> was correlated positively with psychologists' ratings of Expansiveness (.37), Drive toward long-term goals (.41), Talkativeness (.27), Mood swings (.30), Conceit (.38), Elation (.30), Emotional responsiveness (.44), and Concern for the future (.38). It was negatively correlated with ratings of Feelings of inadequacy (-.38), Gastro-intestinal distress (-.31), Weakness of interests (-.40), Lack of adaptiveness to the situation (-.38), Agitation over symptoms (-.30), Lack of faith in self (-.41), and Self-consciousness (-.36).

#### DISCUSSION

Before discussing the significance of these findings for test theory, we should first mention two technical problems of the factor analysis and then discuss the similarity of our findings with those of other factor studies.

The first technical problem concerns the item overlap among scales which will produce spuriously high correlations. The item overlap among CPI scales is not great, but it is large enough to be a matter of concern. There is some evidence that item overlap in inventory tests does not markedly affect the factor structure. Welsh (1956, pp. 246-281) has done factor analyses of MMPI scales with item overlap eliminated and found very similar patterns to those obtained with the full scales. Schutz and Baker (1962, p. 98) stated that "Studies with the MMPI have shown a close correspondence in overall factor pattern of analyses based on scores scored with and without item overlap." Also, Nichols and Beck (1960), in a factor analysis of CPI change scores where item overlap may not be so significant, found factors identical to the

first two factors of this study. It thus appears likely that the factor pattern is not seriously distorted by item overlap, since item overlap is most likely to occur in scales which would be correlated without the influence of common items. Nevertheless, it is likely that the communalities are spuriously high because of the common items.

The second technical problem concerns the method of estimating communalities and the number of factors retained. In this study communalities were estimated as the squared multiple correlation of each scale with all others. Principal axis analysis was used and only three factors were retained for rotation. This was a conservative procedure which produced a minimum number of factors, each accounting for a maximum proportion of variance. Mitchell and Pierce-Jones (1960) used centroid analysis and retained four factors which served to split our third factor into two. Gough<sup>8</sup> in unpublished studies factored the two matrices in the manual with unities for communalities. This tends to result in larger factor loadings, which lead him to retain six factors for rotation, further splitting the factors found in this study. All of these procedures can yield useful information, and the decision among the various alternatives is often arbitrary. For the purpose of constructing factor scales, however, it is desirable for practical reasons to represent only the major clusters by factor scales. The sub-clusters can perhaps be better represented by the individual scales themselves.

The factors of Value Orientation and Person Orientation found in the present analysis are identical to the first two factors in the CPI found by Mitchell and Pierce-Jones (1960) called *Adjustment by Social Conformity* and *Social Poise or Extraversion*, respectively. Our two factors also correspond closely to the first two classes of CPI scales formed by Gough on the basis of early factor and cluster analyses of the CPI scales. They are also very similar to the first two factors found in an analysis of MMPI scales by Kassebaum, Couch, and Slater (1959) called *Ego Weakness versus Ego Strength*

<sup>8</sup> H. G. Gough, personal communication, May 8, 1962.

<sup>7</sup> A shorter version of P can be scored from the 15 scale items which appear in the MMPI. The MMPI booklet numbers and scoring direction for these items are as follows: 2T, 57 T, 343 F, 353 T, 375 F, 391 T, 392 F, 409 T, 415 T, 427 F, 441 T, 455 F, 521 T, 529 T, 547 T.



and *Introversion-Extraversion*, respectively. Eysenck (1953) repeatedly found in his review of factorial studies of personality that the bulk of the common factor variance in these studies was accounted for by a general factor of neuroticism and a bipolar factor of introversion-extraversion. Thus, the two factors found in this study seem to be the same ones indentified in many other analyses. We have preferred the titles Value Orientation and Person Orientation, however, to avoid the confusion which would arise from such commonly used terms as neuroticism, ego-strength, and introversion-extraversion.

We can now return to the problem presented in the introduction of the interpretation of the factors in terms of three possible sources of covariation among scales: (a) common factors among the criteria, (b) common criterion relevant variance among the items not subsumable under (a), and (c) common error variance among the items. The former possibility, common factors among the criteria, is suggested by the similarity of V to the factor of neuroticism or ego weakness and P to the factor of extraversion-introversion found in factor analyses of diverse behavioral measures. However, as Becker (1960) has pointed out, it is dangerous to conclude that inventory factors are identical to behavioral factors solely on the basis of apparent similarity of content. Factor analyses of scales P and V in a matrix of behavioral measures would be necessary to settle this question. Until such evidence becomes available, the latter two possibilities mentioned above must be considered.

It has become common practice for many writers to interpret factors in inventory scales in terms of some response set or response style which is a characteristic of response to a particular item form, rather than representative of factors also found in non-test-behavior. Jackson and Messick (1958) for example, stated that "In light of accumulating evidence, it seems likely that the *major common factors in personality inventories of the true-false or agree-disagree type*, such as the MMPI and the California Psychological Inventory, *are interpretable primarily in terms of style rather than specific item content*" (italics are theirs)

(p. 247). As evidence that acquiescence response style is an important factor, they note that Welsh's MMPI Factor Scale A is composed of items predominantly keyed "true" and Factor Scale R of items all keyed "false." Our Factor Scales P and V do not offer such clear evidence of acquiescence response style. Seventy-eight percent of the items of V are keyed false, but P has about equal numbers of true and false responses. In spite of the predominance of false responses in V, it is presumptuous to conclude that it is correspondingly weighted with acquiescence. Responding direction is related to content in the CPI, and one of the ways one shows positively valued tendencies in this test is by responding "false" to items dealing with impulsivity, hostility, and tension.

Another response style which has been felt to be an important determinant of inventory factors is social desirability. For example, Edwards and Heathers (1962) have pointed out that the factor loadings of MMPI scales on the first factor found by Kassebaum, Couch, and Slater (1959), called ego-strength, correspond closely to the correlations of these same scales with the Social Desirability scale. We suspect that the same is true of our first factor. The shorter version of V scored from the MMPI correlated .63 with the SD scale in the sample of mental patients. As was the case with acquiescence, the interpretation of this finding is difficult because the item content is related to the social desirability of the item. It would be interesting to see correlations of P and V with some of the pure measures of acquiescence and social desirability that have been developed where the response style is not confounded with item content.

Since it is possible that our factor scales, particularly V, have a large response style component, it is important to consider whether the response style is predominantly error which should be discarded or whether it is an important aspect of personality and possibly highly predictive of external criteria. The former possibility appears to be the implication of Edwards and Heathers (1962), while Jackson and Messick (1958) favor the latter. McGee (1962) has reviewed the evi-

dence relating response styles to personality variables and has found it largely inconclusive. The fact that P and V have been shown to be related to a number of non-test behaviors suggests that they are not composed predominantly of irrelevant response styles which should be discarded as errors of measurement. However, on the basis of the present evidence it is not possible to decide between the two remaining possibilities. The factor scales could be either (a) similar to factors in the criteria and dependent mainly on the item content or (b) factors of stylistic response to the item form which are different from the criterion factors, but correlated with them.

The factor scales P and V represent the two major clusters of CPI scales. They do not, however, account for all of the variance in the standard scales nor do they represent any of the scale interactions. Thus, the practitioner who has found the CPI profile meaningful should not consider the factor scales as a substitute for the standard set. He may, however, find the factor scales to be helpful additions to the standard profile, and they may prove to be valuable in the calculation of empirically derived prediction equations.

## REFERENCES

- BECKER, W. C. The matching of behavior rating and questionnaire personality factors, *Psychol. Bull.*, 1960, 57, 201-212.
- EDWARDS, A. L., & HEATHERS, LOUISE B. The first factor of the MMPI: Social desirability or ego strength? *J. consult. Psychol.*, 1962, 26, 99-100.
- EYSENCK, H. J. *The structure of human personality*. New York: Wiley, 1953.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- GOUGH, H. G. The adjective check list as a personality assessment research technique. *Psychol. Rep.*, 1960, 6, 107-122.
- JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, 55, 243-252.
- KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, 23, 226-236.
- LORR, M. Multidimensional scale for rating psychiatric patients. *VA tech. Bull.*, 1953, No. TB 10-507.
- MCGEE, R. K. Response style as a personality variable: By what criterion? *Psychol. Bull.*, 1962, 59, 284-295.
- MITCHELL, J. V., & PIERCE-JONES, J. A factor analysis of Gough's California Psychological Inventory. *J. consult. Psychol.*, 1960, 24, 453-456.
- NICHOLS, R. C., & BECK, K. W. Factors in psychotherapy change. *J. consult. Psychol.*, 1960, 24, 388-399.
- SCHUTZ, R. E., & BAKER, R. L. A factor analysis of the Kuder Preference Record—Occupational, Form D. *Educ. psychol. Measmt.*, 1962, 22, 97-104.
- WEBSTER, H. Maximizing test validity by item selection. *Psychometrika*, 1956, 21, 153-164.
- WELSH, G. S. Factor dimensions A and R. In Welsh, G. S. & Dahlstrom, W. G. (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univ. Minnesota Press, 1956.

(Received August 31, 1962)

## SIBLING PATTERNS AND SOCIAL ADJUSTMENT AMONG NORMAL AND PSYCHIATRICALY DISTURBED SOLDIERS

WILLIAM VOGEL<sup>1</sup> AND CARL G. LAUTERBACH<sup>2</sup>

*Walter Reed General Hospital*

Birth order, and number, age, and sex of siblings were related to social adjustment in a group of 79 psychiatric patient soldiers and a group of 117 normal soldiers, with age of Ss and socioeconomic status held constant. It was found that: (a) Psychiatric patients were more often "only" children than were normal Ss. Character and behavior disorders tended to be "only" children more often than schizophrenics or neurotics. (b) In families of 5 and more siblings, normals more often than patients came from the first and last birth positions. (c) Patients who were judged to have a good prognosis, and married patient and normal Ss, more often than patients with a poor prognosis, and unmarried patient and normal Ss, tended to come from large families. (d) Married patients, and patients who were returned to duty, more often had younger brothers than patients who remained single, or patients who had to be discharged from the service.

There has been considerable interest over the years in the problem of how experience with siblings in childhood affects the psychological and social adjustment which one establishes in adulthood. The purpose of this study was (a) to investigate relationships between certain factors influencing the nature of experience with siblings (birth order, family size, and age and sex of siblings) and incidence and prognosis of psychopathological conditions, and (b) to investigate relationships between these same sibling patterns and marital status.

Spiegel and Bell (1959), in an exhaustive review of the literature on relationships between family size, birth order, and psychiatric diagnosis, report no definitive evidence on how birth order or family size is related to neurosis, psychosis, or character and behavior disorder. They attribute the contradictory nature of the research literature mainly to (a) the absence of control groups or else the use of improperly constituted control groups, (b) the employment of widely differing subject populations in the various studies without consideration of how the special characteristics of the particular popu-

lation used might affect the results of the study, and (c) the use of different criteria of psychopathology in the various studies. Schachter (1959), reviewing the literature independently of Spiegel and Bell, concludes that "the many studies on ordinal position and various indicators of behavior disorder present a pattern of negative results and of weak and inconsistent trends." The literature reported in this area since these reviews has been relatively sparse. However, there have been two studies (Farina, Barry, & Garnezy, 1962; Schooler, 1961) both suggesting that schizophrenics in families of five or more siblings tend to be born in the last half of the birth order. It is unclear what generality this finding might have, since it has not yet been replicated on other diagnostic groups nor has it been related to other possibly relevant factors such as socioeconomic class within the schizophrenic population.

Despite the general lack of consistent empirical evidence in support of expectations that birth order and family size relate to kind and degree of psychopathology, it is possible that these relationships occur in complex interaction with other variables. The present writers concur with Spiegel and Bell (1959) that the nature of the connection between sibling patterns and psychopathology will remain unclear until account

<sup>1</sup> Now at Worcester State Hospital, Worcester, Massachusetts.

<sup>2</sup> Now at United States Military Academy, West Point, New York.



or control is taken of variables which are likely to affect directly the basic conditions in which the subjects were raised. For this reason, in the present study sibling patterns were compared in normal and pathological groups which had been equated for sex, age, education (as an index of socioeconomic class), mean number of siblings, and presence of both natural parents in the home at least until each subject's sixteenth year of life.

The second purpose of this paper, to investigate the relationship between marital status and sibling patterns in normal and psychopathological populations (controlling for the variables enumerated above), leads into an area where few relationships have been reported. Toman (1959) has reported relationships in both normal and psychopathological populations between marital discord and sibling patterns in childhood, although Toman has not worked with the married-unmarried dichotomy which is the focus of interest in this paper.

#### METHOD

**Data collection.** The biographical data were collected from normal and psychiatric subjects by questionnaire. The subjects were asked to list their age, education, marital status, number and sex of older and younger siblings, and to state whether or not they were living with both of their biological parents until they were at least 16 years of age. No questionnaires had to be rejected because of incompleteness, or inability of the subjects to follow instructions. The medical data on the psychiatric group were collected from the patients' final diagnostic summaries. In each case the patient's diagnosis had been concurred in by two or more psychiatric residents or psychiatric staff members.

**Subjects.** The 79 psychiatric patients comprising the psychopathological group and the 117 subjects comprising the normal control group were all American-born, Caucasian male soldiers of enlisted rank who were patients of the Department of Neuropsychiatry at Walter Reed General Hospital. All had been raised by both natural parents until at least 16 years of age. Of the 79 psychiatric patients 24 were officially diagnosed by the psychiatric staff as schizophrenics, 36 as character and behavior disorders, and 19 as neurotics; this admission was the first for any of the patients to a military or civilian neuropsychiatric hospital. Identifying data on the two groups are presented in Table 1.

It was felt that the comparable socioeconomic status of the two groups was assured by their equivalent age, marital status, education, and the fact that all of the subjects were enlisted military

TABLE 1  
BIOGRAPHICAL DATA OF NORMAL AND  
PSYCHIATRIC GROUP

Type of information	Normal subjects (N = 117)	Psychiatric subjects (N = 79)
Mean age	26.6	28.7
Mean number of siblings	4.85	4.41
Education:		
Mean school grade completed	11.8	11.6
% Grammar school or less	9	9
% Some high school	71	77
% Some college	20	10
% College graduate	0	4
Occupation:		
% Unskilled		26
% Semiskilled		21
% Clerical, white-collar, or skilled		51
% Professional		2
% Career military enlisted personnel	37	44
% Married subjects	53	52

Note.—No statistically significant differences in any of the categories exists between the two groups. No occupational data was collected from the normal subjects.

personnel, with about 40% of each group being careerists. The two groups did not differ significantly on any of these variables, nor did they differ on mean number of siblings. The nature of this information, in addition to the large average number of siblings with whom the members of each group were raised, marks both groups as being principally from the lower half of the population socioeconomically.

#### RESULTS

##### *Sibling Patterns of Normal Soldiers and Psychiatric Patients*

The 79 psychiatric patients and 117 normal controls were compared on their own birth order among their brothers and sisters, the number of siblings each subject had, and the age and sex of their siblings. Within the psychiatric population itself, schizophrenics, neurotics, and character and behavior disorders were compared on the same three variables.

**Birth order.** As Spiegel and Bell (1959) report, confusion has resulted because birth order data may be reported in several different ways. It is possible that the subjects'

birth order be reported so that first born, as opposed to second born, or third born, etc. be grouped regardless of the size of the family from which the person came: for example, the second of 2 children and the second of 10 children may both be put in the same category of second-born child. It is also possible to take account of family size; for example, to compare first-born sons from families of three, only to other first-born sons from families of three. Alternatively, one might focus on ordinal position so that first-, intermediate-, and last-born children are the objects of study, regardless of family size. All three methods were employed in this study.

In comparing the 79 psychiatric subjects and 117 normal control subjects on distribution of birth order without regard to family size, the data were arranged in a  $2 \times 6$  contingency table. Five cells represented the first through fifth birth positions and the sixth cell represented the sixth and later birth positions, which were grouped in the sixth cell in order to avoid small expected frequencies. The normal and psychiatric populations did not differ, the percentages in the comparable cells being very similar ( $\chi^2 = 5.74$ ,  $p < .40$ ). The schizophrenic ( $N = 24$ ), character and behavior disorder ( $N = 36$ ), and neurotic groups ( $N = 19$ ) were compared in a  $3 \times 4$  contingency table (the fourth and later birth positions were grouped in the fourth cell because of small expected frequencies in some cells). The results did not approach statistical significance. These data do not allow for the rejection of the null hypothesis of no difference in birth order between psychiatric and normal subjects in this study, nor do they allow the rejection of the null hypothesis of no difference in birth order among the three main psychiatric diagnostic groups, when analyses are done without regard to family size.

When the normal and psychiatric samples were compared for differences in birth order with family size controlled, so that separate analyses were done for families of 2, 3, 4, 5, 6, 7, or 8 members, no significant differences were obtained. Similar results were

found when comparisons were made within the psychiatric population, comparing the three diagnostic groups. Admittedly this was not an adequate test of the null hypothesis of no difference between groups when family size is held constant, because of the fact that the number of subjects coming from a family of any given size was at times quite small. For example, in the  $\chi^2$  test in which the largest number of people were compared (psychiatric patient versus normal soldiers from families of three siblings), only 42 people were involved. As family size increased and the effects of birth order were investigated in families of four or more, the number of persons involved became less, and, of course, the number of contingency cells to fill became greater, so that expected frequencies in particular cells fell below 5.

When the normal and psychiatric samples were compared on ordinal position to see if the samples differed in comparative number of subjects who were first-born, intermediate-, or last-born siblings, no significant differences were obtained. When, however, this analysis was repeated, distinguishing families with four or less siblings from families with five or more siblings as in the studies of Schooler (1961) and Farina, Barry, and Garmezy (1962), it was found that of the 57 normal subjects from families of five or more, 11 normals were first born, 39 normals were intermediate-, and 7 normals were last-born. Of 31 patients from families of five or more, 1 subject was first-born, 29 were intermediate- and 1 subject was last-born ( $\chi^2 = 7.21$ ,  $p < .05$ ). There was no similar effect in families with four or less siblings. This analysis could not be repeated between diagnostic groups because too few patients from families of five or more siblings were born in first or last position.

*Family size.* The normal and psychiatric groups were next compared on the number of siblings with which each subject was raised. The data were arranged in a  $2 \times 7$  contingency table, with the first cell containing "only" children and the seventh cell representing the subjects with six or more siblings, grouped in the seventh cell to avoid small expected frequencies. The two groups



TABLE 2  
NUMBER OF SIBLINGS OF NORMAL AND  
PSYCHIATRIC SUBJECTS

Number of siblings per subject	Percentage of subjects having a given number of siblings	
	Normal group ( <i>N</i> = 117)	Psychiatric group ( <i>N</i> = 79)
0	2	15
1	12	9
2	23	18
3	14	18
4	14	16
5	15	09
6 or more	20	15
Total %	100	100

Note.—Chi square comparing distribution of normal and psychiatric groups, 15.73,  $p < .02$ .

showed a significant difference ( $\chi^2 = 15.73$ ,  $p < .02$ ), the greatest single difference being that 2% of the normals (2 of 117 subjects) were only children, while 15% of the patients (12 of 79 subjects) were only children (see Table 2). This difference is all the more striking if one considers these data in their connection with the data on birth order. While it is true that the normal and psychiatric groups did not differ in their proportions of eldest children (34% of patients were first-born, as compared with 27% of normals), it is also a fact that 12 of the 27 eldest children in the patient group were only children (44%) while just 2 of the 32 eldest children in the normal group were only children (6%;  $\chi^2 = 9.24$ ,  $p < .01$ ).

The three diagnostic groups (schizophrenics, neurotics, character disorders) were compared regarding number of siblings in a  $3 \times 4$  contingency table. (All subjects from families of four or more siblings were grouped in the fourth cell to avoid small expected frequencies.) The findings fell short of statistical significance. What difference there was between these three groups was a function of the fact that only 1 of 24 schizophrenics was an only child, and only 2 of 19 neurotics were only children, but 9 of 36 character and behavior disorders were only children. When these data were arranged in a  $2 \times 3$  contingency table, the

findings fell short of statistical significance ( $\chi^2 = 5.1$ ,  $p < .10$  but  $p > .05$ ).

*Age and sex of siblings.* The normal and psychiatric groups were next compared to see whether they differed in the number of members within each group who had at least one older sister, one younger sister, one older brother, or one younger brother. The comparisons were carried out in a series of  $2 \times 2$  contingency tables (normal versus psychiatric group, number of subjects with one or more older brothers versus subjects with no older brothers, etc.). No significant differences were found, nor were any statistically significant differences found when character disorders, schizophrenics, and neurotics were similarly compared.

### *Prognosis in Psychiatric Patients*

The patients were divided into two groups, a good prognosis group ( $N = 32$ ) whose members the psychiatric staff judged recovered enough to be returned to active military duty (duty group), and a poor prognosis group ( $N = 40$ ) whose members the psychiatric staff caused to be discharged from the army and returned to civilian life as unfit for further military service (discharge group). The duty and discharge groups did not differ in regard to education or occupation; they were of equivalent socioeconomic status. The two patient groups were compared on three aspects of sibling relationships: birth order, the number of siblings each patient had, and the age and sex of their siblings.

*Birth order.* The duty and discharge groups were compared on the birth order variable in a  $2 \times 4$  contingency table. No tendency was seen for the duty group to have a distribution of birth orders different from the discharge group, whether or not family size was taken into account. There was no tendency for the groups to differ in the subjects who were the first-, intermediate-, or last-born among their siblings.

*Family size.* When the duty and discharge groups were compared on the number of siblings each subject had in a  $2 \times 4$  contingency table (the fourth cell representing all those with families containing four or



more siblings), the  $\chi^2$  value was 7.09,  $p < .10$  but  $p > .05$ . What difference there was between the two groups was contributed by a tendency of the duty group to come from families in which there were three or more siblings (including the patient) while the discharge group showed a larger number of "only" children and patients with only one sibling: only 13% of duty patients had one or no siblings, while 35% of discharge patients had one or no siblings.

*Age and sex of siblings.* The duty and discharge groups were compared, in four  $2 \times 2$  contingency tables, to see whether there was any difference between the two groups in the age and sex of their siblings. The members of the duty group more frequently had one or more younger brothers (72% of the duty group had younger brothers while only 37% of the discharge group had younger brothers),  $\chi^2 = 7.10$ ,  $p < .01$ . There was no difference between the two groups in regard to other sibling relationships (older brothers, or older or younger sisters).

#### *Marital Status and Sibling Patterns*

The normal control group contained 54 married and 61 single subjects; the marital status of the two additional normal control subjects was unknown. The psychiatric group contained 30 married subjects, 8 subjects who were divorced or separated from their wives, and 41 single subjects. Examination of the data of the divorced patients showed that regarding the dependent variables with which this study was concerned (birth order, number of siblings with which the subject was raised, age and sex of siblings), the data of the divorced patients were indistinguishable from that of the married patients. Consequently, the patient population was divided into a combined married or once married group (hereafter simply called the "patient married group,"  $N = 38$ ), and a single group ( $N = 41$ ).

In neither the patient nor the normal group did the married and single samples differ on the variable of education. Further, married patients did not differ from the single patients on the occupation variable. It would seem, therefore, that the married normals as compared to the single normals, and the

married patients as compared to the single patients, were of approximately equal socioeconomic status.

*Birth order.* The birth order data for the married and single normal subjects were examined in a  $2 \times 6$  contingency table with the first cell containing single children and the sixth cell all those families with six or more brothers and sisters (including the subject). The  $\chi^2$  was 10.15,  $p < .10$  but  $p > .05$ . What difference existed between the married and single normal subjects was contributed by a tendency of single subjects to be the oldest or next oldest among their siblings. A larger proportion of oldest and next oldest children (66%) were likely to be single, as compared with children who were third or later in birth order (37%). When the similar data for married and once married patients were compared with the data of single patients, however, no differences in birth order were found.

No differences were found when single and married subjects within the patient and normal groups were compared on the birth order variable when family size was taken into account by separately examining birth order effects within each family size (2, 3, 4 siblings etc.). However, the null hypothesis of no differences between groups when family size is held constant is regarded as not having been adequately tested, because of the small number of people from a family of any given size.

Married and single persons in neither the patient nor the normal group tended to differ on having been first, intermediate, or last born in their respective families.

*Family size.* When the married and single normals, and married (or once married) and single patients were compared on the size of the families from which they had come, striking similarities were noted. In the cases of both the normal and patient samples,  $2 \times 6$  contingency tables (the first cell containing only children, the sixth cell containing families with six or more children including the patient) yielded values just short of statistical significance (normal sample,  $\chi^2 = 10.86$ ,  $p < .10$  but  $p > .05$ ; patient sample,  $\chi^2 = 10.17$ ,  $p < .10$  but  $p > .05$ ). In both cases there was a marked tendency for

married persons to come from larger families. When married and single normal persons with three or more siblings were compared with married and single normal persons with two or fewer siblings, it was found that 76% of those from larger families were married compared with 51% of those from smaller families. The comparable figures from the patient population showed 71% of those from larger families were married or were once married, but only 44% of those from smaller families were married or were once married.

*Age and sex of siblings.* When normal married subjects and normal single subjects were compared in four separate  $2 \times 2$  contingency tables to see whether there were any differences between the groups in their having been raised with one or more older brothers, younger brothers, older sisters, or younger sisters (as opposed to their having been raised with no older brothers, younger brothers, older sisters, or younger sisters), there were no significant differences. The only significant difference found when the same comparisons were made for patients was that 71% of married or once married patients had at least one younger brother while only 34% of single patients had a younger brother ( $\chi^2 = 8.61, p < .01$ ).

## DISCUSSION

### *Sibling Patterns of Normal and Psychiatric Patients*

The finding that normal and psychiatric groups differ in that there are seven times as many only children in the psychiatric sample as in the normal control group contrasts with the findings in the literature reviewed by Spiegel and Bell (1959), and suggests that legitimate differences between normal and psychiatric populations have been obscured by failure to control for social class. Holman's (1959) argument as quoted by Spiegel and Bell (1959) that "only" children are more neurotic, but mainly because of greater incidence of broken homes in that group" plainly does not apply to this study, since only the subjects were included in this study who lived with both natural parents throughout childhood. The additional finding that 44% of first-born children in the psychiatric

population were only children, as contrasted with only 6% of first borns in the normal population, puts the data in bolder relief. These data suggest that the only child is subjected to stresses to which children who have siblings are not subjected. It is also possible that the parents of the future psychiatric case had such difficulty raising him that they were discouraged from having additional children. In any event it is interesting to contrast the finding that only children are greatly overrepresented in the psychiatric sample, with the finding that in large families of five or more children, the members of the psychiatric sample are more likely than members of the normal sample to be intermediate children. While intermediate children in large families may be more likely to be neglected than are their first- and last-born siblings, it would also seem true that only children are more likely than other children to be constantly the center of their parents' attention. Either situation apparently is more likely to lead to injury than other sibling patterns. The additional finding that character disorders tend to be only children more often than neurotics or schizophrenics supports several studies cited by Spiegel and Bell (1959) to the effect that only children are overrepresented in the delinquent population, and also indicates that it is the character and behavior group which is primarily responsible for the greater occurrence of only children in the psychiatric as compared with the normal population. These data render some credence to Adler's (1939) theories about the "spoiled only child," which previously have had little firm support in the empirical literature (Spiegel & Bell, 1959).

### *Prognosis in Psychiatric Patients*

The finding that those patients who were returned to duty had younger brothers more frequently than those who did not has implications about the important role that a younger sibling might play in the development of a man's adequacy. It may be that the masculine leadership role in which older brothers who have younger brothers may be cast, is a positive force in the older brothers' development. It is possible, of course, that this latter finding is an artifact due to the



fact that there is a tendency for the duty group to come from larger families, and thus the members of this group would have a greater chance statistically to have a younger brother.

The larger family units from which the duty patients as opposed to the discharged patients tended to come may have rendered the members of the duty group more capable of adjusting to difficult social situations.

### *Marital Status and Sibling Patterns*

It was found that in both the normal and psychiatric samples persons who were married tended more than single persons to come from large families. It was also found that the members of the psychiatric sample who were married were more likely to have younger brothers than were patients who were unmarried. Having a younger brother may incline one to assume a paternal role, and so predispose one to early marriage. Coming from a large family may provide experiences with peers that incline one toward easier acceptance of complex interpersonal situations, such as marriage.

The findings are interesting that both normal and psychiatric subjects who married as opposed to those subjects who remained single, and patients who were judged able to return from duty as opposed to patients who had to be discharged from the service, tended to

come from large rather than small families. Zigler and Phillips (1961) and Lane (1955) found that marriage is a competent index of social effectiveness. Certainly those who were able to return to active military duty must be regarded as more effective persons than those subjects who were so obviously disabled that immediate discharge from the military was necessitated. In each case it would appear that, in this lower socioeconomic class sample, more effective social behavior is associated with coming from a large family.

### REFERENCES

- ADLER, A. *Social interest*. New York: Putnam, 1939.
- FARINA, A., BARRY, H. B. III, & GARMETZ, N. Childhood family structure of schizophrenic patients. Unpublished manuscript. University of Minnesota, 1962.
- LANE, J. Social effectiveness and developmental level. *J. Pers.*, 1955, 23, 274-284.
- SCHACHTER, S. *The psychology of affiliation*. Stanford: Stanford University Press, 1959.
- SCHOOLER, C. Birth order and schizophrenia. *Arch. gen. Psychiat.*, 1961, 4, 91-97.
- SPIEGEL, J. P., & BELL, N. W. The family of the psychiatric patient. In S. Ariete (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1959. Pp. 114-149.
- TOMAN, W. Family constellation as a basic personality determinant. *J. indiv. Psychol.*, 1959, 15, 199-211.
- ZIGLER, E., & PHILLIPS, L. Social competence and outcome in psychiatric disorder. *J. abnorm. soc. Psychol.*, 1961, 63, 264-271.

(Received September 4, 1962)



## SCORES OF PSYCHOTIC PATIENTS ON THE MAUDSLEY PERSONALITY INVENTORY

T. G. CROOKES

*St. John's Hospital, Aylesbury, England<sup>1</sup>*

AND S. J. HUTT

*Park Hospital for Children, Oxford, England*

The hypothesis that psychotics would give normal scores on the N and E scales of the MPI was tested by examining the scores of 40 acute schizophrenics, 20 manics and 20 depressives. The total group did not differ from the normal standardization group on the E scale; it was significantly higher on the N scale, but was considerably nearer to the normal than to a control group of 60 neurotics. The schizophrenics were close to the normal on E, and about half way between normals and neurotics on N. Manics were significantly below normal on N and above normal on E, while depressives were significantly above normal on N and below on E. This finding, that the scale was measuring symptoms rather than a permanent feature of the person, was discussed in relation to other findings. The N scale was not found to make a useful discrimination between psychotics and neurotics, but a low N score plus high E score was found to constitute a manic pattern which distinguished manics well from all other groups.

The Maudsley Personality Inventory (MPI), described by Eysenck (1959a, 1959b), is a short questionnaire designed to measure Neuroticism (N) and Extraversion (E). It consists of 48 questions, 24 for each of the two scales; each answer is scored 2, 1, or 0, so that the range of possible scores on each scale is 0 to 48.

The inventory was administered to a group of psychotic patients for two purposes, one theoretical, one practical. If Eysenck's (1952) theory is correct, that Psychoticism is a dimension which is orthogonal to both Neuroticism and Extraversion-Introversion, then a representative group of psychotic patients should make scores on both MPI scales which approximate, in mean and distribution, to the scores of the normal population. The theoretical aim was to test this hypothesis. The practical aim was to explore the diagnostic possibilities of the inventory. It is a short procedure, and one which most patients find agreeable, so that if it showed good discrimination between any psychiatric groups

it would be a valuable diagnostic aid. For the comparisons required here, scores were also obtained from a group of neurotic and psychopathic patients. The two aims of the investigation are closely related, since the success of the main discrimination, that between neurotics and psychotics, depends on the extent to which the hypothesis about psychotic scores is verified.

### METHOD

In the absence of definite knowledge of the incidence of the different psychoses in the general population, it seemed reasonable to construct our group of equal numbers of the two main "functional" psychotic groups, the schizophrenias and the affective psychoses, and in the affective group to have equal numbers of depressed and manic patients.

The records were collected by giving the inventory to all cooperative patients in St. John's Hospital admission wards who fell into the required diagnostic categories until there were 40 schizophrenics (27 males, 13 females), and 40 affective psychoses, 20 of them manic (12 males, 8 females), 20 depressed (9 males, 11 females). The schizophrenics were all acute cases who had been in hospital less than a year, mostly just a few weeks. At the same time, a group of 60 mixed neurotic and psychopathic cases (35 males, 25 females) was collected, for comparison. The diagnoses were supplied by the psychiatrist in charge of each case. The age range of psychotics was 16-72 (mean 39.5), of neurotics 17-57 (mean 32.9).

<sup>1</sup> Our thanks are due to D. C. Watt, Medical Director of St. John's Hospital, for permission to publish this study and to Deputy Charge Nurses R. Lumsden and K. C. Teed, who administered the test to a number of male patients in their care.

TABLE 1  
MEAN NEUROTICISM AND EXTRAVERSION SCORES

Group	N Score			E Score	
	N	M	SD	M	SD
Normals (Eysenck)	1,800	19.89	11.02	24.91	9.71
Schizophrenics	40	26.35	9.15	22.88	8.36
Affectives	40	19.80	11.23	24.25	9.84
Manics	20	14.60	9.21	28.80	7.66
Depressives	20	25.00	10.82	19.70	9.83
Total psychotics	80	23.08	10.70	23.56	9.10
Neurotics	60	31.35	10.11	—	—

### RESULTS

Table 1 gives the mean scores of each group on each of the two scales. The scores of Eysenck's (1959b) normal standardization group are included for comparison.

The E score of the neurotics is not given, since this depends on the type of neurotics included. Our group was collected without regard to its composition by subgroups, the aim being simply to obtain a group homogeneous with respect to neuroticism.

Table 2 summarizes the significance levels of comparisons between various pairs of groups. All probabilities are two-tailed based on *t* tests. Where a group is compared with Eysenck's normal mean, the latter is treated as the "true" mean of the normal population, and *t* is the comparison of our group mean's standard error with its deviation from the "true" mean. Where two of our groups are compared, *t* is the comparison of the difference between the group means with the standard error of the difference. Comparison of the variances of all the pairs in the Table gives no significant difference. Neurotic comparisons on the E scale are not given, for the reasons mentioned above.

Considered as a whole, the psychotic group does not differ significantly from the normal in E, but is significantly higher in N; it is, however, significantly lower in N than the neurotic group, and considerably nearer to the normals than to the neurotics. The schizophrenics are about midway between normals and neurotics in N, but do not differ significantly from the normal in E. The total affective group does not differ from the

normal on either scale, but taken separately the subgroups differ significantly from each other, and each from the normal on both scales, manics deviating in the direction of low neuroticism and extraversion, depressives in the direction of high neuroticism and introversion. These findings are quite in accord with common sense, but they raise the question of what the scales are measuring. Is it a relatively permanent characteristic of the individual or a temporary one? In other words, are they measuring personality or symptoms? If the conditions are considered as phases in a "cyclic" psychosis, occurring in the "cyclothymic" personality, they must be thought of as occurring in the same type of person, so the scales must be measuring symptoms. But it would also be possible to think of a "manic personality" prone to mania and a "depressive personality" prone to depression. In this case the scales could be reflecting the personality. Two additional operations were carried out to help the decision between these alternatives. (a) The patients had been assigned to their original groups on the basis of the diagnosis of the current admission. It is well known that some manics show quite rapid fluctuations of mood, and the clinical notes of all the manics were examined retrospectively to see if any of them had been actually depressed at the

TABLE 2  
SIGNIFICANCE OF DIFFERENCE (*p*) BETWEEN  
PAIRS OF GROUP MEANS

Groups compared	<i>p</i>	
	N	E
Normals versus psychotics	<.01	>.1
Normals versus schizophrenics	<.001	>.1
Normals versus affectives	>.1	>.1
Normals versus manics	<.02	<.05
Normals versus depressives	<.05	<.05
Neurotics versus all psychotics	<.001	—
Neurotics versus schizophrenics	<.02	—
Neurotics versus affectives	<.001	—
Neurotics versus manics	<.001	—
Neurotics versus depressives	<.02	—
Schizophrenics versus affectives	<.01	>.1
Schizophrenics versus manics	<.001	<.02
Schizophrenics versus depressives	>.1	>.1
Manics versus depressives	<.01	<.01



TABLE 3

DISCRIMINATION OF PSYCHOTIC AND NEUROTIC GROUPS BY N SCALE

Group	Cutoff point					
	Median		Neurotic tenth percentile		Psychotic ninetieth percentile	
	>28	<29	>16	<17	>37	<38
Neurotic (60)	41	19	54	6 (10%)	20 (33%)	40
Psychotic (80)	24	56	57	23 (29%)	8 (10%)	72

time of testing; three definitely, and one probably, were found to have been in a short depressed phase following their manic attack at the time when the inventory was given. The means of these 4 were  $N = 23.8$ ,  $E = 21.5$ , as compared with the means of the remaining 16,  $N = 12.3$ ,  $E = 30.6$ . (b) Those patients were selected from the whole affective group who were definite manic-depressives by the strict criterion of having had actual hospital admissions for both types of attack. There were 9 of these, 3 manic and 6 depressed at the time of testing; means of the 3 manics were  $N = 11.0$ ,  $E = 31.7$ , of the 6 depressed,  $N = 25.3$ ,  $E = 19.7$ . These two operations both support the hypothesis that the scales are measuring a temporary condition or symptoms, rather than permanent personality characteristics. The best way of testing the hypothesis would be to test individual manic-depressives during both types of attack. This we intend to do, but it is clearly a long-term study.

#### *Discrimination of Groups*

The logical employment of the test in an individual case would be to use the N scale to decide between neurotic and psychotic (or neurotic and normal, but this is hardly relevant to clinical practice), then use the E scale to place the individual in the various subcategories of neuroses or psychoses, if indeed there is a useful relationship between E and any of the subcategories. Table 3 shows how the two main groups are divided on the N scale by taking a score as near as possible to the median of the whole group, and by two extreme scores based on the

tenth percentile of the neurotic group and the ninetieth percentile of the psychotic.

When the whole group is divided at the median, 43 patients (31%) are misclassified. Scores below 17 and above 37 have about a 3 to 1 chance of belonging to the appropriate group on the assumption that the population from which they are drawn contains equal numbers of each group, i.e. taking the percentage shown in the table, not the absolute numbers. (For a discussion of the problem of using proportions of this kind when the proportions of the types in the population drawn on are not known, see Crookes, 1961.)

The discrimination of neurotic subgroups by the E scale is discussed elsewhere (Eysenck, 1959a; Foulds, 1961; Sigal, Star, & Franks, 1958), and our experiment was not designed to examine this. Thus no attempt was made to get equal numbers of the subgroups. They actually turned out to be 23 dysthymics (mean  $E = 19.1$ ), 10 hysterics (mean  $E = 27.6$ ), and 27 psychopaths (mean  $E = 26.0$ ). These means are closer together than those of Eysenck's (1959b) corresponding groups, so that discrimination of the three groups will be even more unsuccessful. If the hysterics and psychopaths are combined, and this group compared with the dysthymics, division at the median gives a misclassification of 37% and at the upper end of the scale there is hardly any discrimination. At the lower end, 13 (57%) of the dysthymics have scores below 16, whereas only 2 (5%) of the hysteric-psychopath group have E scores as low as this. So a neurotic giving a very low E score has a good chance of being diagnosed as dysthymic.



It can be seen from Table 1 that the main psychotic groups are not distinguished at all by the E score. The manic subgroup is quite high on E, and this combined with its low N score gives a manic score pattern which is perhaps the only discriminator with any practical value that this study suggests. A score of  $E > 25$ ,  $N < 11$ , is given by 8 out of 20 manics, or 8 out of 16 if the 4 who were actually depressed are excluded, and by only 3 of the remaining 120 patients, 1 depressive, 1 hysteric, and 1 psychopath. This score should be of value in distinguishing mania from schizophrenia, a distinction which is sometimes required in practice. In a recent study by McGuire (1962), in which the MPI was administered to 151 unselected psychiatric in-patients, it was concluded that "the test had no part to play in differential diagnosis, although the N scale might be useful as a screening test." With the exception of its possible value in distinguishing mania and schizophrenia, our own results using the MPI must largely support this conclusion.

#### DISCUSSION

The hypothesis that psychotics would give scores similar to the normal was very nearly confirmed. The psychotic mean on the N scale was significantly higher than the normal, but the difference was only about  $1/3 SD$ . It is worth noting that the 213 normal subjects of Sigal, Star, and Franks (1958) had a mean N score of 23.2, almost exactly the same as our psychotic group. In fact our psychotic group apparently gave less neurotic scores than psychotics in the studies quoted by Eysenck (1960), in which other measures of neuroticism were used. This is probably because we included a considerable group of manics.

The most interesting feature of the study is the scores of the manic-depressives, in which both the scales seem to be indicators of temporary attitudes, or symptoms, rather than permanent characteristics, or personality, although the inventory appears on the face of it to be a measure of personality; subjects are asked to give their "usual way of acting or feeling." This finding ties up with the experiment of Franks, Holden, and

Phillips (1961). They selected those items of the MPI which were the most likely to be objectively observable. They had two groups, one of normal subjects, one of patients (alcoholics), and they asked each person to fill in the scale for himself and for every other member of his group. In each group the members agreed closely in their ratings of the others, but whereas in the normal group the self-ratings also agreed with the ratings by the others, in the patient group the self-ratings showed hardly any such agreement. It seems possible that a scale of this kind measures two different things when used by normals and patients. In the normals it perhaps measures something relatively permanent, and observable, while in some types of patient at least, it seems to measure a temporary subjective attitude. Foulds (1955, 1959) has repeatedly stressed the need to distinguish between "symptom clusters" and "trait clusters," and has developed measures which, in neurotics, are differentially sensitive either to symptom differences or to personality differences. Eysenck himself does not find it necessary to make the distinction. In the *Handbook of Abnormal Psychology* (Eysenck, 1960), for instance, neuroticism is said to be based on a constitutional factor, autonomic balance, but the actual behavior depends on the stress to which the person is subjected; "Outcome = f (Predisposition and Stress) [p. 28]." Neuroticism is presumably that which is measured by tests of neuroticism, since the concept was derived from the intercorrelations of tests, but what do the tests measure, the Outcome or the Predisposition? The answer is suggested by the following passage where Eysenck (1960) is giving a possible reason for the fact that some psychotics were found to give high scores on neuroticism:

... we might regard it (psychoticism) as essentially a powerful stress which, when multiplied by an average predisposition, produces as the final outcome a strongly emotional and neurotic behaviour pattern. In other words, although the person high on psychoticism is not necessarily high on neuroticism, the strong stress produced by his disordered thought processes results in a high  $w \rightarrow k$  sequence leading to strongly emotional reactions [p. 28].

It is clear that he has here abandoned the attempt to use the word "neuroticism" in a single sense. The psychotics in question are high on neuroticism as measured by the tests, in terms of which the word was originally defined; they also show neurotic behavior, but they are said to be not necessarily high on another neuroticism, which must be the underlying constitutional factor. This makes it fairly clear that the tests are thought of as measuring the temporary condition or symptoms rather than a permanent condition, and our findings with manic-depressives would support this as far as the MPI is concerned. For the psychologist working in a practical, therapeutic situation, it is important to know which kind of neuroticism a given test is measuring, if he wants to assess changes due to treatment, or if he wishes to know what aspects of behavior are likely to be amenable to treatment. Hence the importance of the work done by Foulds. It would also be useful to make a similar distinction in the case of measures of psychoticism. Manic-depressives would seem a particularly suitable group for an enquiry into this, and the authors hope to use patients of this type for an investigation into the temporary and permanent characteristics of psychotics.

## REFERENCES

- CROOKES, T. G. Wechsler's deterioration ratio in clinical practice. *J. consult. Psychol.*, 1961, 25, 234-238.
- EYSENCK, H. J. *The scientific study of personality*. London: Routledge and Kegan Paul, 1952.
- EYSENCK, H. J. The differentiation between normal and various neurotic groups on the Maudsley Personality Inventory. *Brit. J. Psychol.*, 1959, 50, 176-177. (a)
- EYSENCK, H. J. *Manual of the Maudsley Personality Inventory*. London: Univer. London Press, 1959. (b)
- EYSENCK, H. J. (Ed.) *Handbook of abnormal psychology*. London: Pitman, 1960.
- FOULDS, G. A. Psychiatric syndromes and personality types. *Psychol. Forsch.*, 1955, 25, 65-78.
- FOULDS, G. A. The relative stability of personality measures compared with diagnostic measures. *J. ment. Sci.*, 1959, 105, 783-787.
- FOULDS, G. A. The logical impossibility of using hysterics and dysthymics as criterion groups in the study of introversion and extraversion. *Brit. J. Psychol.*, 1961, 52, 385-387.
- FRANKS, C. M., HOLDEN, E. A., & PHILLIPS, M. Eysenck's "stratification" theory and the questionnaire method of measuring personality. *J. clin. Psychol.*, 1961, 17, 248-253.
- MCGUIRE, R. J. A study of the M. P. I. used with psychiatric in-patients. *Bull. Brit. Psychol. Soc.*, 1962, 47, 56. (Abstract)
- SIGAL, J. J., STAR, K. H., & FRANKS, C. M. Hysterics and dysthymics as criterion groups in the study of introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, 57, 143-148.

(Received September 4, 1962)

## EFFECTS OF ANGER AROUSAL AND SIMILARITY UPON THE ATTRIBUTION OF HOSTILITY TO PICTORIAL STIMULI<sup>1</sup>

SEYMOUR FESHBACH

*University of Pennsylvania*

ROBERT D. SINGER

*Stanford University*

AND NORMA FESHBACH

*University of Pennsylvania*

A study investigating the role of similarity between the perceiver and the stimulus in determining the projection of experimentally aroused hostile affect was initially carried out with a group of 57 college Ss, and then, due to the ambiguity of the data, was replicated on a sample of 65 college Ss. Prior to and immediately after administration of the anger-arousing or control treatment, Ss judged a series of stick figures and a series of photographs of two-person situations in which the age of a male figure was systematically varied. Both male and female experimental Ss in comparison to control Ss manifested reliable increments in attribution of hostility to college-age male stimulus persons only.

The present study is part of a more general research effort aimed at investigating the effects of emotional arousal on the process of interpersonal judgment. The focal point of these investigations is an assessment of the variables which facilitate the attribution of the perceiver's affective state to some stimulus person. In this sense, these studies can be said to fall within the rubric of "projection." However, as has been stated elsewhere (Feshbach & Singer, 1957; Murstein & Pryer, 1959), the concept of projection encompasses a number of diverse processes and the use of a single label may obscure important functional differences. In accounting for the influence of emotional arousal on the perception of other people, the authors have stressed the role of stimulus generalization and the egocentric orientation which affective involvement is likely to produce.

Two previous experiments (Feshbach & Singer, 1957; Singer & Feshbach, 1962) were concerned with the attribution of the emotion of fear or anxiety. In these studies, evidence was obtained that under conditions of fear

arousal, both normal and psychotic subjects tend to attribute a greater degree of fear to others than under more relaxed conditions. Another group of experiments by Feshbach and Feshbach (1963) based upon Murray's (1933) well-known investigation of the influence of fear upon the attribution of maliciousness also dealt with the influence of fear upon social judgment but with a primary interest in the type rather than degree of distortion resulting from fear arousal. These studies indicated that frightened boys tend to attribute fear to boys their own age, but, at the same time, attribute maliciousness to older men. Their judgments of girls similar to their own age and of older women were relatively unaffected by their emotional state.

It is apparent, then, that the process mediating affective attribution cannot be fully understood by an analysis only of the perceiver's internal state. One also has to consider characteristics of the person being judged and his relationship to the perceiver. The principal thesis of the present study is that similarity between the perceiver and the stimulus person facilitates attribution of affect. Similarity, as is attested by the literature on substitution and stimulus

<sup>1</sup> This study was supported by Research Grant M3213 from the National Institutes of Health, United States Public Health Service.



generalization, is deceptive in its apparent simplicity as an independent variable. Similarity can vary along many dimensions and it is essential, where human beings are concerned, to sample more than one dimension in evaluating the significance of this variable. The particular experimental design employed in this study includes variation of the stimulus object with respect to age and human similarity to the perceiver. Hostility was selected as the affective state to be aroused in order to extend the generality of the previous findings in the area of fear and anxiety.

The function of this design has been considered within the context of an analysis of the role of the emotional state of the perceiver in interpersonal perception. The experimental data are also relevant to the interpretation of projective techniques, especially such measures as the TAT and the Rosenzweig Picture Frustration Test, which employ human interaction situations as the mirror to reflect other characteristics of the respondent. There are a number of studies which have recently been reviewed by Murstein (1961) in which some aspect of similarity such as race and sex between the stimulus object and the respondent has been varied. The use in many of these studies of such indices of projection as story length greatly limits the inferences that can be drawn with respect to the relationship of the material elicited to attributes of the respondent. An advantage of the present design in this regard lies in the ability to specify the attribute—namely, anger and hostility.

## METHOD

### Subjects

The initial sample consisted of 25 control and 32 experimental subjects drawn from introductory psychology classes at the University of Pennsylvania. Due to sizable initial differences between the experimental and control subjects which rendered the results ambiguous, the experiment was repeated using 20 control and 45 experimental subjects who were taking summer session courses in psychology at the City College of New York. Subjects were randomly assigned to either the experimental or the control condition, with a greater proportion being assigned to the former treatment. The distribution of males and females was also unequal. The Pennsylvania control group consisted

of 8 males and 17 females and the experimental group of 16 males and 16 females. The New York control group was comprised of 12 males and 8 females and the experimental group of 23 males and 22 females. In both groups, the average age was about 19.5 and the mean number of years of education was approximately 14.

### Materials

Two classes of stimuli were used. The first type consisted of slides of animated stick figure drawings and the second of slides of TAT type photographs. There were 26 stick figures, chosen from the Sarbin (1954) series. These are simple stick drawings, neutral in sex and suggesting a person in various poses. Those stick figures were selected which could be perceived as engaged in some sort of hostile activity. Information regarding the perceptual properties of these figures was available from previously obtained norms. In addition, prior to the experiment proper, the full series was administered, with five alternative adjective choices provided for each stick figure, to a group of 65 students at the University of Pennsylvania who were asked to select the adjective that best described the figure. The stick figures were then divided into two matched series of 13 slides on the basis of the number of hostile attributions elicited by each stimulus.

These same 65 subjects were also presented 28 TAT-like photographs of two-person interaction situations. These pictorial situations were somewhat similar to the Rosenzweig Picture Frustration Test in that they were so posed that, for each stimulus, one person, labeled A, could be perceived as frustrating a second person, labeled B. The source of frustration varied and included a college-age male, a college age female, an older man, and an older woman. The object of frustration, Person B, was either a college age male or a 12-year-old boy. The task of the subjects was simply to indicate for each interaction how Person B felt. In contrast to the stick figure series, the question was open-ended in that no alternatives were provided.

Two matched series of pictorial situations were then constructed on the basis of the number of hostile attributions elicited by each situation. Series 1 consisted of nine pictures. In five, Person B, the object of frustration, was a college age male and in four he was a 12-year-old boy. Series 2 consisted of pictures similar in sex-age composition to those in Series 1 with respect to both the object and source of frustration, except that the number of same age and younger age stimulus persons were reversed. Ideally, it would have been desirable to use exactly corresponding situations for the younger boy-college age male variation. However, apart from the problem of whether the meaning of the situation would remain the same with a difference in age of the object of frustration, the use of identical pairs of situations introduces possible order effects and, more important, makes transparent the purpose of the variation. In order to minimize the possible confounding of situational

TABLE 1

MEAN CHANGE SCORES IN AGGRESSIVE ATTRIBUTION AS A FUNCTION OF INITIAL LEVEL OF JUDGMENT AND EXPERIMENTAL TREATMENT: EXPERIMENT I

	Stick figures			Pictorial situation		
	Initial score			Initial score		
Experimental <i>M</i>	0, 1 +2.1 (9f, 3m)	2 +4 (4f, 5m)	3, 4 -5 (3f, 8m)	0, 1 +1.8 (4f, 7m)	2 +.8 (8f, 4m)	3, 4 -2 (4f, 5m)
Control <i>M</i>	+ .7 (3f, 0m)	+ .8 (6f, 3m)	-.8 (8f, 5m)	+1.4 (5f, 0m)	+ .2 (9f, 2m)	-.7 (3f, 6m)

Note.—The figures in parentheses represent the distribution of men (m) and women (f) in the respective cell.

differences with age variation, the content of the situations was varied, a wide range of frequencies of aggressive attributions to both younger and college age males was employed, and the sources of frustration for each age group were quite similar in age and sex.<sup>2</sup>

### Procedure

Series 1 of the stick figures and photographs were presented by means of slides to experimental and control subjects who were seen in groups of approximately 10. In order to judge the stick figures, the subjects were provided with an answer sheet which had five alternative adjective choices for each stick figure. In each instance, one of the alternatives was an adjective denoting some hostile feeling or activity. Examples of such hostile adjectives include "annoyed," "angry," "sullen," "mad," and "indignant." For the pictorial situations the subjects were provided with a blank sheet on which they could indicate, for each picture, how they thought Person B felt. The experimental subjects were then subjected to a series of insulting comments by a 29-year-old male experimenter who disparaged their motivation and cooperativeness. His accusations that they were not trying and were apparently not interested in research, etc. were arbitrary and unprovoked. The effectiveness of the procedure in arousing hostility has been demonstrated in previous studies (Feshbach, 1961). The experimenter delivered these comments in a hostile and condescending manner. In the control condition, the same experimenter made some neutral comments concerning the slide projector and the screen.

In both the experimental and control conditions, the subjects then proceeded to judge the second series of stick figures and pictorial situations. These

procedures were carried out in an identical manner in the replication of the experiment, the stick figures always being administered before the pictorial situations.

### Scoring

For both the stick figures and the pictorial situations, one point was given for every hostile ascription to the stimulus. The total scores on each stick figure series could then range from 0 to 13, while the scores on the pictorial situation series could range from 0 to 9. In the case of the latter, a reliable scoring scheme, based upon the initial pre-experimental data, was readily established, there being 95% agreement between two independent scorers. The following words and phrases are typical examples of descriptions of the feelings of the stimulus figures that were scored as hostile: mad, angry, irritated, hostile, irritable, sore, about to lose his temper, about to blow up, and aggressive.

### RESULTS

As was previously noted, the first experiment could not be easily interpreted because of initial differences between the experimental and control subjects. Since the range of possible change was quite restricted, especially in the case of the pictorial situations, one would expect, on statistical grounds, an inverse relationship between initial level and degree of change. This relationship is evident in Table 1. An analysis of the stick figure data presented in Table 1 reflected a significantly higher proportion of experimental females increasing in aggressive attributions as compared to the control females. However, as Table 1 indicates, this difference may be simply a consequence of the differences in distribution of initial scores on this measure between the experimental and control females. It can be seen from Table 1 that

<sup>2</sup> A complete set of these slides can be obtained from the authors. The following examples are cited to indicate the types of situations that were utilized: (a) A college age male is holding a tie and looking at a second college age male, the caption reading, "You don't mind if I borrow your tie?" (b) An older woman is looking at a younger boy while he is opening a door. The caption reads, "Are you going out again?"



TABLE 2

MEAN FREQUENCY OF AGGRESSIVE ATTRIBUTIONS TO STICK FIGURES AND PICTORIAL SITUATIONS BEFORE AND AFTER THE EXPERIMENTAL AND CONTROL TREATMENT:  
EXPERIMENT II

EXPERIMENT 11					
	N	Stick figures		Pictorial situations	
		Before	After	Before	After
Males					
Experimental	23	2.6	2.7	1.7	2.6
Control	12	2.7	2.5	1.7	2.0
Females					
Experimental	22	2.9	2.5	1.8	3.1
Control	18	2.5	3.2	2.0	1.5

with initial level held constant, there is a consistent tendency for the experimental subjects to attribute more aggression than the controls to the stimulus persons in the photographs. While this effect is statistically significant only for males judging same age figures, initially higher scores in judgments of same age figures for the experimental males and initially lower scores for the experimental females in comparison to their respective controls may be confounding factors.

Consequently the experiment was replicated on a new sample. In this instance, random sampling worked as it should, and as can be seen from Table 2, the initial judgments of the experimental and control groups were quite similar. As in the earlier study, change scores in the amount of hostility attributed to the second series of stick figures and pictorial situation stimuli as compared to the initial series were computed for each subject. The distribution of these scores was dichotomized according to those subjects who manifested an increment in hostility on the second series versus subjects who gave fewer or the same number of hostile descriptions to the second series as to the first. It is evident from Table 3 that there is no reliable change in attribution of hostility to the stick figures. However, as Table 4 indicates, both the female as well as the male experimental subjects manifest a significant increment in hostility attribution to the same age male stimuli. Again, as in the initial experiment, there was no significant change in response to the younger age

TABLE 3

FREQUENCY OF POSITIVE AND NEGATIVE CHANGES IN NUMBER OF AGGRESSIVE ATTRIBUTIONS TO STICK FIGURES AS A FUNCTION OF SEX OF PERCEIVER:  
EXPERIMENT II

	Males		Females	
	+	0-	+	0-
Experimental	9	14	8	14
Control	3	9	5	3
	$p > .10$		$p > .10$	

stimulus persons for either the male or female subjects.

### DISCUSSION

The results indicate that angered individuals will project hostile affect. The results also indicate that the stimulus object plays a critical role in the attribution process. One significant property of a stimulus is its cue value with respect to the relative frequency with which a particular interpretation is given to the stimulus (Birney, 1958; Haber & Alpert, 1958). The present experiment has been concerned with the effect upon projection of another property of the stimulus—namely, similarity to the perceiver. Although all three categories of stimuli—the stick figures, the same age males and younger age boys—had been selected on the basis of having aggressive value, responses to only the same age males discriminated between the anger aroused and the control groups.

A theoretical model which attempts to explain this form of projection as a joint

TABLE 4

FREQUENCY OF POSITIVE AND NEGATIVE CHANGES IN NUMBER OF AGGRESSIVE ATTRIBUTIONS TO THE STIMULUS PERSON AS A FUNCTION OF HIS AGE AND THE SEX OF THE PERCEIVER: EXPERIMENT II

	Same age				Younger age			
	Males		Females		Males		Females	
	+	0-	+	0-	+	0-	+	0-
Experimental	18	5	16	6	8	15	8	14
Control	3	9	2	6	4	8	3	5
	$p^a = .005$		$p^a = .0027$		$p^a > .10$		$p^a > .10$	

<sup>a</sup>  $p$  values (one-tailed) determined by Fisher's exact test.



function of a lowered threshold for responding with motive-related imagery and the presence of stimulus cues capable of evoking such imagery would not adequately account for the obtained data. Similarity between the stimulus person and the perceiver appears to be a critical variable in determining whether the perceiver will attribute his own emotional state to the stimulus person. Judgments of the stick figure stimuli which are symbolic representations of humans but which deviate markedly in physical similarity to "real people" were unaffected by the experimentally induced affective state of the perceiver. In addition, the experimental subjects failed to attribute or "project" their anger onto stimulus persons who were considerably younger than they. However, it is noteworthy that the female subjects in the experimental group as well as the male subjects projected their irritation and anger onto the male stimulus persons. While one cannot completely assess the effects of differences between the stimulus and the perceiver without varying the sex of the stimulus, it is evident that the presence of similarity with respect to the sex dimension did not vitiate projection by the female subjects.

The present theoretical formulation of the process mediating affective attribution assigns a significant role to stimulus generalization which in turn is presumed to be dependent upon stimulus similarity. Although "similarity" is a fundamental psychological variable which is postulated to mediate phenomena ranging from transfer of training (Hovland, 1951) to displacement (Miller, 1944) its effects from a priori standpoint are often difficult to assess. The relevance of a similarity dimension may vary with the situation and the particular response under study. Theoretical formulations are required which will lead to predictions as to the conditions under which particular dimensions of similarity become pertinent. By exploring the identifications and perceived similarities of the college-age student, for example, one may be able to provide a basis for predicting when variations in sex and variations in age will influence generalization of some specific behavior.

As a further complication, if there is con-

siderable anxiety in connection with the expression of a motive, then one might expect that judgments of dissimilar rather than similar stimuli would be more revealing. The authors in a number of unpublished experiments have been unable to demonstrate reliably that the induction of aggression-anxiety produces a relative shift in attribution of anger to dissimilar stimulus persons. Whatever the ultimate status of this particular hypothesis, it is evident that the dimensions of similarity require more systematic analysis and specification if similarity is to be used as a predictive variable.

# REFERENCES

- BIRNEY, R. C. Thematic content and the cue characteristics of pictures. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton: Van Nostrand, 1958. Pp. 630-643.
- FESHBACH, S. The stimulating versus cathartic effects of a vicarious aggressive activity. *J. abnorm. soc. Psychol.*, 1961, 63, 381-386.
- FESHBACH, S., & FESHBACH, NORMA. Influence of the stimulus object upon the complementary and supplementary projection of fear. *J. abnorm. soc. Psychol.*, 1963, 66, 498-502.
- FESHBACH, S., & SINGER, R. Effects of fear arousal and suppression of fear upon social perception. *J. abnorm. soc. Psychol.*, 1957, 55, 283-288.
- HABER, N. H., & ALPERT, R. The role of situation and picture cues in projective measurement of the achievement motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton: Van Nostrand, 1958. Pp. 644-663.
- HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 613-689.
- MILLER, N. E. Experimental studies in conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 431-435.
- MURRAY, H. A. The effect of fear on the estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, 4, 310-329.
- MURSTEIN, B. I. The role of the stimulus in the manifestation of fantasy. In J. Kagan and G. E. Lesser (Eds.), *Contemporary issues in thematic apperceptive methods*. Springfield: Charles C. Thomas, 1961. Pp. 229-273.
- MURSTEIN, B. I., & PRYER, R. S. The concept of projection: A review. *Psychol. Bull.*, 1959, 56, 353-374.
- SARBIN, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology*. Reading, Mass.: Addison-Wesley, 1954. Pp. 223-258.
- SINGER, R., & FESHBACH, S. Effects of anxiety arousal in psychotics and normals upon the perception of anxiety in others. *J. Pers.*, 1962, 30, 574-587.

(Received September 19, 1962)

## SIMULATION ON THE CALIFORNIA PSYCHOLOGICAL INVENTORY AND THE ADJUSTMENT OF THE SIMULATOR

FRANCIS M. CANTER

Patton State Hospital

This is a study of the relationship between ability to present a good picture on the California Psychological Inventory and the actual life adjustment of S. A group of 50 alcoholics and 50 applicants for Ward Aide positions, both highly motivated to appear "normal," were compared. Additionally, a group of well-functioning Aides, a group of Better Adjusted Alcoholics, and a group of Poorer Adjusted Alcoholics were compared under "fake good" instructions. Results indicate a positive relationship between ability to present a good picture, under standard or faking conditions, and the relative adjustment of S. The results are related to Gough's concepts of socialization and it is concluded that these questionnaire responses are meaningful aspects of the personality and not simply sets to be corrected for.

Most studies on the susceptibility of questionnaire-type personality tests to deliberate distortion by the subject have used college students or other relatively well-functioning people as subjects. They have usually been asked to take the test under the standard "honest" conditions and then again under instructions to distort the picture in some specified way, with comparison of the score differences as indication of the degree to which the test is vulnerable to "faking." In the *Manual for the California Psychological Inventory* (Gough, 1957) there is evidence that subjects can change their scores on this test to a considerable extent, although this usually resulted in tell-tale changes in the Wb, Gi, or Cm scales. Some tests attempt to correct for these possibilities of "faking" by the addition of validating scales, e.g., the K and L scales of the MMPI, or by forcing a choice between two items of previously determined equal social desirability (Edwards, 1957).

Relatively little attention has been paid to the actual personal adjustment of the subject or his own "natural" inclinations to present himself in the best possible light. In a recent study (Dicken, 1960) a relationship was found between ability to simulate a "good" profile on the California Psychological Inventory (CPI) and the sophistication and technical knowledge of the subject. Grayson and Olinger (1957), using psychi-

atric patients as subjects, found that some of their patients under "fake good" instructions could change their MMPI profiles in certain ways and that ability to do this was related to favorability of prognosis for discharge from the hospital. These studies employed the technique of giving instructions to fake good with subjects who presumably had no special motivation to answer other than honestly under the standard instructions. A point of possible theoretical and practical importance is the ability of subjects to present a "good" picture under standard instructions, where there is presumed strong motivation to present such a picture and the relationship of this ability to the subject's actual life adjustment. It is specifically suggested that falsification of test scores and the capacity to do so may be a personality variable of considerable importance in its own right rather than merely an undesirable and incidental factor to be "corrected for."

The present study compared the CPI responses of two groups, each group presumably strongly motivated to present themselves in a favorable light, one group being objectively "well-adjusted" and the other "poorly-adjusted." Additionally, the performance of the poorly-adjusted group under fake good instructions was examined in relationship to rated degree of adjustment and was compared with the well-adjusted group's performance under fake good instructions.



## METHOD

### *Subjects*

The subjects of the study consisted of a group of 50 male alcoholic patients in a state hospital and a group of 50 male applicants for the position of aide at this same hospital. The alcoholic patients (poorly adjusted) were all involuntary admissions who were committed to the hospital for an indefinite period. The applicants, who were accepted for employment and worked for periods of at least 6 months and usually much longer, were rated periodically on their effectiveness and personal adjustment. Only those with a satisfactory rating after at least 6 months were included in the well-adjusted group. In other respects, the groups did not differ significantly from each other. The applicants and the patients came from the same general socioeconomic backgrounds with occupations in the farming, small tradesman, salesman, factory worker, or construction labor areas. Median age for the alcoholic group was 41 and was 43 for the applicant group. Median years of school was, for both groups, 11. Mean Raven Progressive Matrices score was at the sixtieth percentile for the alcoholics and the sixty-third percentile for the applicants.

### *Procedure*

The applicants were examined with the California Psychological Inventory and the Raven Progressive Matrices prior to acceptance for employment. They were given to understand that employment depended in part upon "satisfactory performance" on the psychological tests. It is assumed that these applicants were highly motivated to present a favorable picture on the California Psychological Inventory.

The alcoholic patients were examined within 5 days after admission (their first admission to this hospital) with a battery of psychological tests including the California Psychological Inventory and the Raven Matrices. These patients were involuntary and quite anxious to be discharged from the hospital as soon as possible. They did not regard themselves as "mental patients," felt insulted or put upon by their commitment to a mental hospital, and emphasized, in individual interviews and groups meetings, their basic normality and lack of difference from other people except for their bad habit of drinking too much. The psychological examination was regarded by these patients as an effort to find out whether they were "crazy" or not, particularly since the test battery was prefaced with the Rorschach. In general, it may be said that these patients regarded themselves, at least consciously, as personally "normal" except for the isolated area of drinking (which they were inclined to attribute to their biochemistry) and had ample motivation to convince the hospital staff of their eligibility for discharge at the earliest possible date.

Of this group of 50 patients, 30 who were still available were requested to take the CPI again with instructions to

imagine you are applying for a job you really want and your employer will judge from this test whether to hire you or not. Answer the test in such a way as to give the best possible impression of yourself.

These 30 patients were then rated for general adjustment and acceptance of their drinking problem according to this scale:

1. Past history: maximum score of 8 points with 1 point each for items relating to the patient's work stability, maintenance of family relationships, no previous hospitalization for drinking, freedom from legal difficulties, and definite efforts to handle drinking by resort to Antabuse, Alcoholics Anonymous, or psychiatric help
2. Locally devised questionnaire: maximum score of 4 points, based upon 34 items dealing with present acceptance of drinking problem
3. Interview: maximum score of 4 points, given for acceptance of self as personally maladjusted with drinking as a symptom, 2 points for considering self as "normal" but with a drinking problem, and no points for regarding self as normal and with no drinking problem

Scores on this rating of relative adjustment ranged from 2 to 16 with a median score of 9.5. The 15 subjects in the upper half were compared with the 15 subjects in the lower half. These two groups did not differ significantly on such variables as age, education, Raven score, or average scores on the standard administration of the CPI.

Seventeen of the applicants who were available were also asked to take the CPI under the fake good instructions. Because of circumstances, these tests were obtained after the applicant had been hired.

## RESULTS AND DISCUSSION

In Table 1 are presented the individual scale average raw scores of the alcoholics as a group and the applicants as a group under standard instructions. The differences between the two groups are highly significant on all scales except the Cm scale. In Table 2 are presented the individual scale average raw scores of the Applicants, the Better Adjusted Alcoholics, and the Poorer Adjusted Alcoholics (designated as Groups A, B, and C, respectively) under the fake good instructions. All groups are able to improve their scores but the Applicants show more improvement than the Better Adjusted Alcoholics and these, in turn, show more improvement than the Poorer Adjusted Alcoholics. Although there are fewer statistically significant differences between groups under the faking condition, the general trend of the scores is consistent.



TABLE 1

CALIFORNIA PSYCHOLOGICAL INVENTORY RAW SCORES  
OF ALCOHOLIC AND APPLICANT GROUPS:  
NORMAL ADMINISTRATION

Scale	Alcoholic (N = 50)		Applicant (N = 50)		<i>t</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Do	21.9	5.0	32.1	5.9	9.18**
Cs	14.8	4.5	20.8	4.4	6.70**
Sy	20.8	5.3	26.9	3.7	6.57**
Sp	31.0	6.1	35.6	5.8	3.83**
Sa	17.4	3.9	22.3	4.5	5.77**
Wb	34.9	8.4	40.6	4.2	4.22**
Re	26.1	4.1	32.9	4.6	7.76**
So	29.3	6.1	39.2	4.3	8.50**
Sc	26.8	7.9	34.7	6.3	5.48**
To	16.9	4.8	24.6	5.2	7.62**
Gi	14.3	6.0	23.8	6.8	7.31**
Cm	26.4	1.5	26.6	1.7	0.62
Ac	23.3	4.7	31.0	4.8	8.03**
Ai	14.5	3.9	19.5	5.0	5.52**
Ie	32.7	5.3	41.0	5.9	7.36**
Py	8.4	2.6	11.6	2.8	5.86**
Fx	6.7	3.6	8.5	4.8	2.11*
Fe	16.3	5.6	17.2	4.8	0.85

\*  $p < .05$ .  
\*\*  $p < .01$ .

These results suggest that, under standard conditions of administration and where there is presumed motivation on the part of subjects to present themselves as normal or well adjusted, there is a positive relationship between the ability to present a good picture on the CPI and the actual life adjustment of the subject. In addition, under fake good instructions, subjects are able to improve their test pictures to a considerable degree but even here there is limitation by actual life adjustment. In this connection, it is noted that the Gi scores of all groups under the faking conditions show the expected and usually found peak, indicating a conscious dissembling. However, Gough (1957) has pointed out that individuals obtaining high Gi scores do, in fact, tend to create a better impression on others and show some realistically better adjustment.

A central issue here is possibly the question of what is meant by dishonesty in answering questionnaire items. (Probably, the results of the standard administration are more intrinsically meaningful in this study than the

faking conditions, because instructions to fake something arouse a number of factors which are difficult to assess, including the subject's interpretation of the purpose of the procedure, etc.) As mentioned above, most of the studies on response set, social desirability, and conscious faking have used subjects who by conventional criteria are fairly adequate and "socialized" people who at least have a background of personal socialization and experience to help them to fake good. Gough (1948, 1960a, 1960b) has made a significant and central point of the idea that socialized behavior is rooted in the role-taking experiences of development and the asocial or inadequate person lacks the set of interactional expectancies and reaction patterns which permit more socially adequate behavior (including answering questionnaire items on the relevant behavior). This is not contradictory to the findings of Jackson (1960) who concluded that response set and social desirability factors account for much of the variance in the CPI if the subjects in such investigations are individuals from the same cultural matrix with the same "folk"

TABLE 2

CALIFORNIA PSYCHOLOGICAL INVENTORY SCORES  
OF AIDES AND PATIENTS UNDER INSTRUCTIONS  
TO FAKE GOOD

Scale	Group A (N = 17)		Group B (N = 15)		Group C (N = 15)		<i>t</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	A-B	B-C
Do	37.2	4.3	35.7	2.0	31.0	9.3	1.25	0.73
Cs	26.8	3.1	23.6	3.0	20.5	4.9	2.88**	1.31
Sy	33.5	4.1	29.4	3.1	26.8	3.9	3.12**	1.46
Sp	40.5	3.8	35.2	4.2	34.9	3.9	3.60**	0.13
Sa	23.7	3.0	22.9	2.8	21.5	2.3	0.74	1.43
Wb	42.8	2.1	41.7	1.8	37.8	4.4	1.54	2.44*
Re	37.7	3.7	35.7	2.2	28.1	6.3	1.83	2.40*
So	43.5	3.1	42.7	2.7	34.3	7.1	0.75	2.04*
Sc	42.0	4.3	39.1	4.2	29.0	7.7	1.87	1.84
To	28.8	3.2	26.2	2.9	18.8	4.6	2.34*	3.65**
Gi	32.5	5.2	29.9	3.4	23.9	5.2	1.64	2.14*
Cm	26.7	1.4	26.5	1.1	25.9	3.1	0.44	0.38
Ac	32.3	4.0	33.5	1.9	26.5	5.6	1.07	2.80**
Ai	21.2	3.0	18.9	3.0	13.9	3.4	2.09*	3.40**
Ie	43.7	3.7	42.4	3.0	35.3	8.6	1.06	1.20
Py	12.9	2.0	12.3	1.6	9.7	2.2	0.90	4.82**
Fx	7.6	2.4	4.7	2.0	3.1	3.0	3.62**	1.72
Fe	14.7	3.8	15.3	4.1	14.9	3.9	0.41	0.13

\*  $p < .05$ .

\*\*  $p < .01$ .

Note.—Group A = Psychiatric aides; Group B = Better adjusted alcoholic patients; Group C = Poorer adjusted alcoholic patients.

concepts and hence are able readily and naturally to fit into the way of thinking which is reflected as good adjustment on the CPI. Consistent with this is the finding by Sundberg and Bachelis (1956) that their subjects were better able to fake bad than fake good on the California F Scale and the Gough *Pr* scale. From some recent research, Dicken<sup>1</sup> has indicated that efforts to "correct" for response bias on the scales of the CPI resulted in little improvement or even a reduction in validity, suggesting that normal scores, response set and all, are meaningful aspects of the "real" personality, just as the ability to fake good says something meaningful about the person's present or potential adjustment, given the appropriate situation. Thus, a person who can and naturally does answer items in a more socialized way is drawing upon an actual and potential store of possibility for good adjustment than the person who has to concentrate on answering the items "correctly" and this person in turn has more possibility for good adjustment than the person who naturally does not or

cannot answer in a socialized manner, even when asked to bend his conscious attention to it.

#### REFERENCES

- DICKEN, C. F. Simulated patterns on the California Psychological Inventory. *J. counsel. Psychol.*, 1960, 7, 24-31.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- GOUGH, H. G. A sociological theory of psychopathy. *Amer. J. Sociol.*, 1948, 53, 359-366.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto: Consulting Psychologists Press, 1957.
- GOUGH, H. G. Cross-cultural studies of the socialization continuum. *Amer. Psychologist*, 1960, 15, 410-411. (Abstract) (a)
- GOUGH, H. G. Theory and measurement of socialization. *J. consult. Psychol.*, 1960, 24, 23-30. (b)
- GRAYSON, N. M., & OLINGER, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, 21, 73-77.
- JACKSON, D. N. Stylistic response determinants in the California Psychological Inventory. *Educ. psychol. Measmt.*, 1960, 20, 339-346.
- SUNDBERG, N. D., & BACHELIS, W. D. The fakability of two measures of prejudice: The California F Scale and Gough's *Pr* scale. *J. abnorm. soc. Psychol.*, 1956, 52, 140-142.

<sup>1</sup>C. Dicken, personal communication, September 1962.

(Received September 20, 1962)

## EFFECT OF PROFESSIONAL TRAINING AND AMOUNT AND ACCURACY OF INFORMATION ON BEHAVIORAL PREDICTION<sup>1</sup>

JANIS H. WEISS<sup>2</sup>

*University of Iowa*

Behavioral prediction tests were administered to 60 PhD clinical psychologists and 60 PhD physical scientists. Some judges had only minimal descriptive material from which to predict while others had this plus a typescript of a self-interview to utilize. In half the cases, the minimal information was not accurately descriptive of the S to be predicted. When only minimal information was available to the judge, psychologists performed significantly more accurately than physical scientists. Physical scientists, however, perform more accurately than psychologists when a greater amount of information is available. Use of the self-interview led to improvement in the prediction of physical scientists but did not significantly affect the accuracy of psychologists' predictions.

With the increasing interest in clinical psychology have come numerous studies investigating variables that influence how clinicians arrive at a valid set of predictions about a client's behavior. The assumption is usually made that the formal training and professional experience of a clinician are important factors in increasing the accuracy of his predictions of client behavior. Although most clinical training programs are firmly rooted in this premise, the empirical findings do not always support it.

Typically when the predictive performance of graduate students in psychology is compared with the performance of PhD psychologists, no significant differences in accuracy are found (Cline, 1955; Grigg, 1958; Kessen, 1957; Soskin, 1954a, 1959). Only Klehr (1949) was able to report that experienced clinicians fared better than graduate students in predictive accuracy. In studies which compare psychologically trained groups to groups of judges who are trained in other areas, usually physical scientists (Cline, 1955; Luft, 1950; Taft, 1955), it has often been found

that the latter groups perform more accurately than do psychologists. The results of Cline's study in which engineering trainees were used are an exception. Indeed, Taft (1955) summarized the literature by suggesting that physical scientists and possibly other non-psychologists may be more capable of judging others accurately than are either psychology students or clinical psychologists. The fact that groups of judges of similar age and academic achievement have not been utilized in the above studies suggested a need for additional research.

Another variable which is assumed to affect accuracy of behavioral prediction is the amount of information made available to the judge. Studies by Kostlan (1954), Sines (1959), Cline (1955), Borke and Fiske (1957), and Dailey (1952) indicate that accuracy of prediction can be enhanced by giving the judge more information from which to predict. However, several interesting limitations seem to apply to this general conclusion. One is that prediction appears to be enhanced only by some type of interview or case history data, not by psychological test data. Secondly, an increase in accuracy of prediction over that made from minimal information may be found only when the judge is supplied immediately with the total amount of data to be studied. Contrawise, Gage (1953) and Soskin (1954a, 1959) have reported results which indicate that

<sup>1</sup> This study is based upon a dissertation presented to the Department of Psychology of the University of Iowa in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

<sup>2</sup> The author wishes to express her appreciation to Leonard D. Goodstein for his guidance throughout the course of this study and to the Old Gold Development Fund for its financial aid.



increased amounts of information do not lead to more accurate predictions and in some cases to less accurate ones. Other investigators (Cline & Richards, 1960; Dymond, 1953; Hathaway, 1956; Leventhal, 1957) have pointed out that a small amount of data (usually class membership information such as age, sex, and vocation) accounted for nearly all the predictive accuracy of judges and that additional information appeared not to enhance predictions appreciably. Though accuracy of information is certainly presumed to affect predictive accuracy, few attempts have been made to experimentally manipulate the correctness of the information. A preliminary study by Harvey (1938) apparently indicated that incorrect information is limited in its effect upon prediction of behavior when contradictory evidence is also present.

The purpose of this study was to determine the effect of type of professional training, amount of information, and accuracy of information upon the prediction of the actual responses of individuals in real-life social situations. By using a group of males with PhD degrees in one of the physical sciences and PhD psychologists, it was possible to investigate the effect of training on prediction of behavior while presumably holding age, intelligence, and level of academic achievement constant. Another purpose of the study was to investigate the effects upon predictive accuracy of differing amounts of information. To do this some judges had only information on the sex, age, and student status of the subject to be predicted while others had this information plus a typescript of a self-interview given by the subject. Finally, half of the judges were given accurate minimal information as to the presumed psychological health of the subject, i.e., that the subject was a volunteer, hired to participate in this study, while the others were given inaccurate information, i.e., that the subject was receiving psychotherapy in the University Counseling Service for his (her) personal problems.

Because of the equivocal results of the research reported above, the following predictions were based for the most part on common sense expectations:

*Hypothesis 1.* PhD psychologists should be more accurate predictors of the behavior of individual college students than PhD physical scientists.

*Hypothesis 2.* Within these two professional groups those judges who are given the greatest amount of information about the subject to be judged should be better predictors than judges receiving less information.

*Hypothesis 3.* Judges receiving the correct information about the psychological health of the subject should predict the subject's behavior more accurately than those receiving incorrect information. However, the incorrect information may produce poorer predictions only for those judges who have the least amount of other information.

## METHOD

### *Experimental Material*

Three subjects, Ray, Ron, and Ruth, were hired to participate in an extensive program of psychological testing and interviewing. The subjects were naive volunteers from an undergraduate psychology class. As part of the test battery, these students were placed in a room with a tape recorder and asked to talk about themselves for no longer than 30 minutes. The format for this self-interview was developed by Gilbert (1959).<sup>3</sup> The subjects were given 10 topics about which to talk which were intended to cover in brief form such things as their educational and vocational goals, family experiences, emotional relationships, ethical and religious views, and so on. Typescripts were made of these complete self-interviews in as exact detail as possible.

### *Criterion Tests*

The criterion tests are similar to those described by Soskin (1954a). Each of the subjects to be used in this study was interviewed by an experienced clinical psychologist in an attempt to get information about their childhood, adolescent, and college experiences. Further, two close friends of each subject, one male and one female, were also interviewed for 1 hour each in an attempt not only to verify information given by the subject but also to obtain independent assessments of the subject. For each of these subjects a multiple-choice, four-foil, 28-item test was constructed, the test being based on actual behavior as reported by the subject and, whenever possible, verified by the friends. A sample question is as follows:

There was a lot of talk at the restaurant where Ruth works about something she had done. The incident involved:

<sup>3</sup> W. M. Gilbert and J. McV. Hunt, personal communication, 1959.

- a. Ruth's necking with a Negro employee at a party.
- b. Ruth's organizing a strike for higher wages.
- c. Ruth's reporting to work one night drunk.
- d. Ruth's keeping tips left for other waitresses.

The judges used in the development and item analysis of the three tests were all undergraduate volunteers from the Introductory Psychology class. The first two drafts of the tests were given to 80 and 50 judges, respectively. Items were revised if the answers given by the judges to each one of four foils of an item were not distributed within plus or minus three standard deviations from the expected chance mean. This means the tests were constructed so that any response biases for naive judges were minimized and so that the information yielded by the stems alone did not enhance accuracy. The accuracy of the answers given by the judges was predetermined on the basis of the information from the subjects' and friends' interviews. In the final draft, given to 132 judges, the mean correct score on each of the three tests was as follows: The Ron Test, 8.02; The Ray Test, 7.51; and The Ruth Test, 7.03. Upon retest of 80 judges (1 month interval) the mean correct score on each test was: The Ron Test, 8.01; The Ray Test, 7.30; and the Ruth Test, 7.26. None of these changes was significant. For a more complete discussion of the test development see Thorson (1962).

### Judges

Sixty professional clinical psychologists with a PhD degree and affiliation with either the Divisions of Clinical and/or Counseling Psychology of the American Psychological Association constituted the psychologist group. Their names were selected from the 1961 edition of the *American Psychological Association Directory*. The second group consisted of 60 physical scientists with a PhD degree in either physics, chemistry, or engineering. Their names were selected from Volume I, tenth edition of *American Men of Science*. As far as could be determined, all but four of the psychologists were presently connected with an academic institution and all but six of the physical scientists worked in an industrial concern. All judges were males who were born after 1900 and who had received their undergraduate and graduate degrees in the United States. In order to obtain these judges, 120 psychologists and 120 physical scientists were contacted by mail. Of the original sample contacted, 49 of the psychologists and 56 of the physical scientists completed the task. Because these judges did not constitute the necessary number of judges needed and did not distribute evenly over all cells, it was necessary to contact 21 additional psychologists and 12 additional physical scientists.

### Procedure

Prior to the mailing of the materials, each of the primary professional groups was further subdivided

on a random basis into 12 subgroups of five judges each. The divisions were made on the basis of the presence or absence of the self-interview typescript, the accuracy of the minimal information, and the subject to be predicted. The minimal information was included in the main body of the instructions given to all judges. Half of the judges were told that the subject to be predicted is a (age)-year-old (sex) college student who was hired to participate in this study (accurate information). The other judges were told their subject is a (age)-year-old (sex) college student who is presently being seen in the University Counseling Service for help with his (her) personal problems (inaccurate information). In addition a complete typescript of the subject's self-interview was included in the packet sent to half of the judges. Instructions were the same as those for the minimal information group except that these judges were asked to read the self-interview prior to taking the criterion test. Since there were three different subjects to be predicted, one third of the judges received the criterion test pertaining to Ruth, one third received the one pertaining to Ray, and one third received the one pertaining to Ron.

### RESULTS

The mean accuracy scores and standard deviations for the treatment combinations are included in Table 1. In order to assess the differences in predictive accuracy between these groups, a  $2 \times 2 \times 2 \times 3$  factorial analysis of variance (Lindquist, 1956) was used, the factors being type of professional training (P), presence or absence of the self-interview information (I), correctness of the minimal information (M), and subject to be predicted (T). The results of this analysis indicated that only the  $P \times I$  interaction was significant ( $p < .001$ ). It is suggested, therefore, that the effect of increased information depends on the type of professional training of the judge. From Table 1 it can be noted that this interaction is due to the superiority of the physical scientists over the psychologists in the self-interview condition but superiority of the psychologists over the physical scientists in the minimal information condition. None of the interactions or main effects involving accuracy of the minimal information and/or subject to be predicted was significant.

Factorial analyses ( $2 \times 2 \times 3$ ) were then run for psychologist and physical scientist groups separately, in order to investigate the effect of information while holding profes-



TABLE 1  
MEANS AND STANDARD DEVIATIONS OF THE ACCURACY SCORES FOR ALL  
TREATMENT COMBINATIONS

		Self-interview					
		Correct minimal			Incorrect minimal		
		Ruth	Ray	Ron	Ruth	Ray	Ron
Psychologist	<i>M</i>	6.80	7.20	6.80	6.60	5.40	7.60
	<i>SD</i>	1.72	0.75	1.94	1.50	1.02	1.62
Physical scientist	<i>M</i>	9.00	9.20	9.40	6.40	7.60	9.60
	<i>SD</i>	2.68	2.79	1.74	1.20	1.36	2.15
		No-self-interview					
		Correct minimal			Incorrect minimal		
		Ruth	Ray	Ron	Ruth	Ray	Ron
Psychologist	<i>M</i>	6.00	7.80	8.00	8.80	7.00	8.20
	<i>SD</i>	1.79	1.17	2.45	3.54	2.28	2.14
Physical scientist	<i>M</i>	6.40	6.80	6.80	5.40	5.60	7.40
	<i>SD</i>	1.50	0.98	0.75	2.06	0.80	2.24

sional training constant. For the psychologists, none of the interactions or main effects was significant though the main effect of the self-interview information approached significance ( $.10 > p > .05$ ). It should be noted that the direction of the difference was negative, i.e., the use of the self-interview tended to lead to poorer predictions. Within the physical scientist group, the main effect of the increase information condition was significant ( $p < .001$ ). The direction of this difference was positive, indicating that use of the self-interview led to improvement in predictive performance for the physical scientists. Additional factorial analyses were run in order to assess the effect of professional training while holding amount of information constant. Within the self-interview information condition, the only significant result was the effect of professional training ( $p < .001$ ) with the physical scientists performing *more* accurately than psychologists. In the minimal information condition the main effect of professional training was again significant ( $p < .05$ ) with the psychologists performing more accurately than the physical scientists.

## DISCUSSION

It would appear that psychologists are better predictors of the behavior of college students than are physical scientists only when minimal information is available. Though it may be that psychologists have developed more accurate knowledge of the behavioral base rates in the college student population than have physical scientists, the amount and intensity of association with college students that the judges in each group have had in recent years must be taken into account. The physical scientists were purposely selected so as to minimize their association with college students while psychologists were selected who had maximum association with students. It remains for further research to investigate whether there may be personality factors associated with the choice of work setting which may be related to interest in and perhaps ability to predict the behavior of college students.

When the self-interview information is utilized in making predictions, physical scientists are found to be more accurate pre-



dictors than are psychologists. It would appear, therefore, that physical scientists are better able to utilize the information contained within the self-interview in order to enhance their predictive accuracy. Soskin (1954a, 1954b, 1959) has suggested that psychologists tend to overestimate a subject's maladjustive propensities when using test data and thus tend to decrease their predictive accuracy on multiple-choice instruments by choosing foils more indicative of maladjustment. Hathaway (1956) and Cronbach (1955) have further suggested that psychologists are too concerned with the uniqueness of the individual to be predicted and tend to neglect their conceptions of the behavior of the average person in any particular population. Certainly in most clinical training programs primary emphasis is placed on increasing the clinician's sensitivity to signs of maladjustment and to aspects of the personality which differentiate a particular individual from others. It may be that the physical scientist does not alter his frame of reference toward the maladjustive or unique end of the continuum and thus utilizes the available material more accurately, particularly when the individual to be predicted belongs to a "normal" population.

Perhaps the finding of major importance in terms of the criterion tests is that although improvement in scores is possible, the tests are quite difficult and all the means hover around the expected chance mean. Though there is some evidence from an undergraduate population that on retest judges tend to respond to the items in a manner consistent with their original responses (Thorson, 1962) and may not be responding on a random basis, the reader still may question whether the criterion tests afford an adequate or meaningful opportunity for the psychologist to test his predictive abilities. No attempt has been made to construct items which would be judged by psychologists as relevant information though most of the items appear to have adequate face validity. Since there is no criterion available by which material may be assessed as to its predictive importance and its relevance to the functioning of a psychologist

at work, the experimenter is, of necessity, left to his own judgmental devices. Construction and choice of criterion measure would appear to be an area in which much important research still remains to be done.

It was expected from prior research (Jentsch, 1958) that judges would differ in the accuracy of their predictions depending upon which subject they were asked to predict. Of more interest is the lack of interaction between the subject to be predicted and any of the other variables. This means that though the amount of information influenced the accuracy of prediction, the ease with which the subjects could be predicted remained essentially the same. Coupled with Jentsch's findings, it makes more feasible the notion that only a limited number of subjects need to be used in research such as this while still allowing for generalizability of results to a meaningful larger population.

#### REFERENCES

- BORKE, HELEN, & FISCHE, D. W. Factors influencing the prediction of behavior from a diagnostic interview. *J. consult. Psychol.*, 1957, 21, 78-80.
- CLINE, V. B. Ability to judge personality assessed with a stress interview and sound-film technique. *J. abnorm. soc. Psychol.*, 1955, 50, 183-187.
- CLINE, V. B., & RICHARDS, J. M., JR. The accuracy of interpersonal perception: A general trait? *J. abnorm. soc. Psychol.*, 1960, 60, 1-7.
- CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, 52, 177-193.
- DAILEY, C. A. The effects of premature conclusions upon the acquisition of understanding of a person. *J. Psychol.*, 1952, 33, 133-152.
- DYMOND, ROSALIND E. Can clinicians predict individual behavior? *J. Pers.*, 1953, 22, 151-161.
- GAGE, N. L. Explorations in understanding others. *Educ. psychol. Measmt.*, 1953, 13, 14-26.
- GRIGG, A. E. Experience of clinicians and speech characteristics and statements of clients as variables in clinical judgment. *J. consult. Psychol.*, 1958, 22, 315-319.
- HARVEY, S. M. A preliminary investigation of the interview. *Brit. J. Psychol.*, 1938, 28, 263-287.
- HATHAWAY, S. R. Clinical intuition and inferential accuracy. *J. Pers.*, 1956, 24, 223-250.
- JENTSCH, R. Influence of psychological-mindedness and information on accuracy of prediction. Unpublished doctoral dissertation, State University of Iowa, 1958.
- KESSEN, W. The role of experience in judging children's photographs. *J. abnorm. soc. Psychol.*, 1957, 54, 375-379.

- KLEHR, H. Clinical intuition and test scores as a basis for diagnosis. *J. consult. Psychol.*, 1949, 13, 34-38.
- KOSTLAN, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, 18, 83-88.
- LEVENTHAL, H. Cognitive processes and interpersonal predictions. *J. abnorm. soc. Psychol.*, 1957, 55, 176-180.
- LINDQUIST, E. F. *Design and analysis of experiments*. Houghton Mifflin: Boston, 1956.
- LUFT, J. Implicit hypotheses and clinical prediction. *J. abnorm. soc. Psychol.*, 1950, 45, 756-759.
- SINES, L. K. The relative contribution of four kinds of data to accuracy in personality assessment. *J. consult. Psychol.*, 1959, 23, 483-492.
- SOSKIN, W. F. Bias in postdiction from projective tests. *J. abnorm. soc. Psychol.*, 1954, 49, 69-74. (a)
- SOSKIN, W. F. Frames of reference in personality assessment. *J. clin. Psychol.*, 1954, 10, 107-114. (b)
- SOSKIN, W. F. Influence of four types of data on diagnostic conceptualization in psychological testing. *J. abnorm. soc. Psychol.*, 1959, 58, 69-78.
- TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, 52, 1-23.
- THORSON, MARGARET. The development of several criteria for assessing changes in the prediction of human behavior. Unpublished master's thesis, State University of Iowa, 1962.

(Received October 3, 1962)

## NOTES AND COMMENT

### APPLICABILITY OF THE NORMS OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN TO FIVE-YEAR-OLDS

MICHAEL BELDOCH<sup>1</sup>

*Teachers College, Columbia University*

The WISC was administered to 90 children between the ages of 5-0 and 6-11, 29 of whom were between the ages of 5-0 through 5-3, in accordance with the selection procedures of a school for gifted children. The youngest children were found to have a mean Full Scale IQ approximately 10 points lower than that of the group 5-4 in age and older. A retest 11 months later on 14 of the children of the youngest age group and 15 of the older Ss showed a rise in IQ scores of 9 points for the youngest group, those of the children originally tested when 5-4 or older remaining unchanged. The results of the study were interpreted as demonstrating a need for revision of the extrapolated norms now provided in the WISC manual for the age range 5-0 through 5-3. A question was raised as to the usefulness of the WISC with 5-year-old children.

The use of the WISC with subjects at the lower age ranges of the test has been criticized by a number of investigators. Delp (1953) noted evidence that the difficulty of the test for children in the lower age brackets may make results incomparable with those of other instruments. In comparing the Stanford-Binet Intelligence Scale with the WISC, Krugman, Justman, Wrightstone, & Krugman (1951) found that the younger the child, the greater the difference in scores in favor of the S-B relative to the Verbal and Full Scale IQ scores on the WISC.

When standardizing the WISC, Wechsler (1949) tested his sample of normal children within 1.5 months of their midyear and did not include subjects of 5 years, 0 months through 5 years, 3 months of age. Nonetheless, the test manual provides tables of normalized scale scores for every 4-month interval between the ages of 5 and 15 years. Anderson (1953) has written that

it is difficult to see how Wechsler can arrive at norms for every 4-month period and be consistent with his earlier objection to interpolation and extrapolation of scores and his insistence that a person's results be compared with norms based upon results of testing persons of his own age [p. 478].

Referring to the lowest age groups, especially the duller members of the group, Fraser (1959)

maintained that the standardization sample of WISC performances had not been adequate.

The present study attempted to test the clinical impression that the use of the WISC with children 5 years, 0 months through 5 years, 3 months of age, norms for whom are included in the WISC manual but are not based on standardization data, put them at a distinct test disadvantage when compared with older subjects.

#### METHOD

In the Guidance Laboratory of Teachers College, Columbia University, a psychological clinic, almost 200 children were administered the WISC in a program of selection of students for a municipally operated school for the gifted in New York City. Of this group, 90 were between the ages of 5-0 and 6-11 at the time of testing. All testing was done by graduate students in psychology.

TABLE 1  
IQ SCORES ON FULL SCALE OF THE WISC FOR CHILDREN  
5-0 THROUGH 6-11

Ages	N	M	SD	t
5-0-5-3	29	111.03	15.23	-1.68*
5-4-5-7	19	122.00	10.13	
5-8-5-11	18	120.22	13.53	
6-0-6-11	24	121.00	14.92	

\*  $p < .025$ .

<sup>1</sup> Now at the College of the City of New York.



TABLE 2

RETEST IQ SCORES ON FULL SCALE OF THE WISC FOR  
CHILDREN 5-0 THROUGH 6-11

Age at original testing	N	M	SD	t
5-0-5-3	14	121.57	13.20	-.43
5-4-6-11	15	123.33	7.97	

Table 1 shows the mean IQ scores and standard deviations of the Full Scale of the WISC for the sample of 90 children, grouped into 4-month intervals for ages 5-0 through 5-11 and in a year-interval for the 6-year-olds. It will be noted that the youngest age group (5-0 through 5-3) was significantly lower in mean Full Scale IQ score than the combined three older groups ( $p = .025$ ), the three older groups having been combined because of highly similar mean IQs. In order to determine if this difference was due to the age factor or arose through some special sampling error, it was decided to test the children once again after the lapse of 11 months.<sup>2</sup> If the youngest children had been placed at a test disadvantage, the passage of time should reveal a sharp rise in the mean IQ of this group while those of the older children should remain relatively constant.

Approximately 1 year later, it was possible to contact 29 of the children and reschedule them for testing. Of this retest group, 14 had been 5-0 through 5-3 at the time of the original testing while 15 had been 5-4 through 6-11. The graduate students who administered this test were different from the original testers and had no knowledge of the nature of the study, the children involved, or their previously obtained IQ scores.

### RESULTS

Table 2 shows the mean IQ scores and standard deviations on the Full Scale of the WISC for each of the two retested groups. From these results it can be seen that the younger children now obtain scores very similar to those of the older subjects and that the previously noted differences have almost entirely disappeared. The scores of the older children remained virtually

<sup>2</sup> The writer is indebted to Rosalea Schonbar for this suggestion.

the same. *t* tests comparing the samples retested with the rest of the original samples show no significant differences to obtain between the smaller retest groups and the rest of the *N* of 90 from which they were drawn.

### DISCUSSION

It seems clear that the extrapolated norms for ages 5-0 through 5-3 available in the WISC manual are in need of revision since they tend to place children of that group at a distinct test disadvantage. The results of this study would question the advisability of the use of the WISC with children of this age, in confirmation of Delp (1953). Since even among children of essentially superior intellectual ability, the subjects of this investigation, the WISC scores were spuriously low, it might be argued that with youngsters of lesser capacity, even greater difficulty will be encountered. The subjective impressions of the psychologists who examined the subjects in the original testing were that the younger children were often unable to understand the directions printed in the manual and were "at a loss" with the test materials. It was thought that the major difficulty had been experienced on the Performance Subtests, but an analysis of the Verbal Scale versus the Performance Scale IQ scores failed to support this hypothesis.

### REFERENCES

- ANDERSON, J. M. Review of the Wechsler Intelligence Scale for Children. In O. K. Buros (Ed.), *Fourth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1953. Pp. 477-479.
- DELP, H. A. Review of the Wechsler Intelligence Scale for Children. In O. K. Buros (Ed.), *Fourth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1953. Pp. 479-480.
- FRASER, ELIZABETH D. Review of the Wechsler Intelligence Scale for Children. In O. K. Buros (Ed.), *Fifth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1959. Pp. 558-559.
- KRUGMAN, JUDITH I., JUSTMAN, J., WRIGHTSTONE, J. W., & KRUGMAN, M. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for children. *J. consult. Psychol.*, 1951, 15, 475-483.
- WECHSLER, D. *Wechsler Intelligence Scale for Children: Manual*. New York: Psychological Corporation, 1949.

(Received May 28, 1962)

## RESEARCH CREATIVITY IN PSYCHOLOGY GRADUATE STUDENTS

MARTHA T. MEDNICK<sup>1</sup>

*University of Michigan*

The validity of the Remote Associates Test (RAT) was investigated. Graduate student advisors rated individuals on a research creativity check list. These ratings were intercorrelated with RAT and Miller Analogy Test scores and with grade point averages. The ratings correlated significantly only with RAT scores ( $r = .55$ ,  $p < .005$  for 43 Ss). This supports the use of this test as a selection device where research creativity is a desired trait.

Within an associative framework, creative thinking is regarded as the process of forming new combinations of associative elements. The more mutually remote the elements utilized, the more creative is the solution. An additional stipulation of this definition is that the solution be useful or meet specified requirements. The associative approach, which has been described in greater detail by Mednick (1962) has led to the development of a measure of the creative thinking process called the Remote Associates Test (RAT) (Mednick, 1962; Mednick & Mednick, 1962). The test is composed of 30 three word items of the following type:

Example 1:    rat            blue    cottage

Example 2:    railroad    girl    class<sup>2</sup>

The words are chosen from mutually remote associative clusters and the task is to find a mediating link between them. The link is strictly associative and does not follow rules of logic, concept formation, or problem solving. Considerable normative data has been gathered on various forms of this test.

Mednick (1962) reported that when faculty advisors were asked to judge whether their psychology graduate students were high or low in research creativity, the resulting judgments related highly positively to the students' RAT scores. The criterion measure used was rough

and did not, of course, give any notion of the value of the contribution of the middle range of scores. This study represents an attempt to assess this relationship in a more detailed fashion.

### METHOD

#### *Creativity Rating Scale (CRS)*

In the course of his work on scientific creativity and productivity, Taylor (1957) developed a Thurstone-type research creativity rating scale. For our own work, 24 of the items were reworded so as to make them suitable for the rating of graduate students.<sup>3</sup> The scale values used are identical to those used by Taylor. When these scales were distributed, the research advisor was asked to rate only those individuals whom he knew well.

#### *Remote Associates Test*

The RAT was administered to graduate students entering the Psychology Department at the University of Michigan in 1959 and again in 1960; a group of Northwestern University graduate students was also tested. This gave us a total of 81 Michigan and Northwestern students. At the time of the ratings, 23 of these students had terminated their graduate education and 15 others had not worked intensively enough with any one research advisor to be rated in a reliable manner. CRS ratings were obtained on the remaining 43 students. The ratings were gathered between 1 and 2 years after RAT administration and the raters were not aware of the RAT scores achieved by their advisees. Miller Analogy Test (MAT) scores and grade point averages (GPA) were obtained from subject's records but were not available on the entire group.

### RESULTS AND DISCUSSION

The means, ranges, and standard deviations of the RAT scores of various graduate student sub-

<sup>3</sup> Copies of the revised scale may be obtained from the author.

<sup>1</sup> This research is supported by the Cooperative Research Program of the Office of Education, Contract No. 1073, Martha T. and Sarnoff A. Mednick, coinvestigators.

Now Visiting Research Associate, Psychological Institute, Kommunehospitalet, Copenhagen, Denmark.

<sup>2</sup> Answers to sample RAT items: 1. cheese; 2. working.

groups are given in Table 1. The norms for the two undergraduate groups were also included in order to provide some comparison data. It can readily be seen that the specially selected group of honors students did better than unselected undergraduates and that the graduates as a whole obtain scores similar to the former group. The characteristics of the group which received ratings appear to be quite similar to those of the graduate student group as a whole.

The average CRS scores ranged from 18.0 to 56.2 with a mean of 40.7. The internal reliability of the CRS was checked in the following manner. The items were arranged in order of their scale values. These were then divided into odd and even halves and two separate CRS values were computed for each individual. The Spearman-Brown reliability coefficient was .93.

The correlations which were obtained between the various measures are presented in Table 2. We should note that the mean MAT score of this group was 71.1 with a standard deviation of 12.12. The norm group mean was 65.8 with a standard deviation of 13.5 (Miller, 1960). The effect of this restriction of the distribution was evaluated by a correction of the correlation of the MAT and RAT (Guilford, 1956). The correlation of the two tests was thus raised to .44. This indicates that the distribution of the MAT scores in our group is not different enough from those of the original norm group to have an important effect on the correlations reported here.

The correlation of greatest interest from our

TABLE 1  
REMOTE ASSOCIATES TEST SCORES OF PSYCHOLOGY  
GRADUATE AND UNDER-GRADUATE STUDENTS

Group	N	M	Range	SD
Psychology graduate students				
Rated	43	19.8	10-27	4.3
Drop-outs	23	20.8	09-28	4.9
Not rated	15	20.0	16-26	3.1
Total	81	20.1	09-28	4.0
Undergraduate students				
Freshmen	312	16.3	00-28	5.2
Psychology honors	23	20.0	11-28	4.2

TABLE 2  
MATRIX OF INTERCORRELATIONS OF  
RATING SCALE AND TEST SCORES

Measure	RAT	GPA	MAT
CRS			
r	.55**	.06	-.08
N	43	16	24
RAT			
r		-.11	.41*
N		26	25
GPA			
r			-.03
N			25

\*  $p < .025$ .

\*\*  $p < .005$ .

point of view is, of course, that of the RAT and the CRS. It is evidently possible to predict judgments of graduate student research performance from these test scores. This particular type of rating was not significantly related to MAT scores or to GPA. This property of the test may recommend it for use in selection of individuals in those instances where research creativity such as is measured by the CRS is a desired trait.

#### REFERENCES

- GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.
- MEDNICK, S. A. The associative basis of the creative process. *Psychol. Rev.*, 1962, **69**, 220-232.
- MEDNICK, S. A., & MEDNICK, M. T. A theory and test of creativity. In, *Proceedings of the XIV International Congress of Applied Psychology*. Vol. 5. *Industrial and business psychology*. Copenhagen: Munksgaard, 1962. Pp. 40-47.
- MILLER, W. S. *Manual, Miller Analogies Test*. New York: Psychological Corporation, 1960.
- TAYLOR, D. W. Variables related to creativity and productivity among men in two research laboratories. In C. W. Taylor (Ed.), *Research Conference on the Identification of Creative Scientific Talent*. Salt Lake City: Univer. Utah Press, 1957. Pp. 20-54.

(Received July 23, 1962)



## MMPI PROFILES OF EPILEPTICS:

### A FURTHER EVALUATION

EDWARD J. JORDAN, JR.<sup>1</sup>

*Research Service, Medfield State Hospital, Massachusetts*

This study investigated (a) the validity of MMPI-derived epilepsy signs, (b) whether the signs may reflect epilepsy specifically or brain disorder in general, and (c) whether age and/or IQ affect incidence of signs. 32 epileptics were compared with 19 neurological patients, 19 psychotics, and 26 neurotics, all male veterans, in percentage of patients surpassing given numbers of signs. Optimal cutoff points failed to discriminate epileptics from any group but psychotics, and this was due to unique characteristics of the psychotic group. The signs are thus not regarded as generally valid for either epilepsy or brain disorder. Age and verbal IQ apparently do not affect incidence of signs.

Hovey, Kooi, and Thomas (1959) report a set of 14 signs in MMPI profiles which discriminated epileptic from psychiatric patients beyond the .001 confidence level. A cutting point of seven or more signs correctly identified 67% of the epileptics in the sample and misclassified as epileptic only 18% of psychiatric patients.

The present study is primarily an attempt to replicate the findings of Hovey, et al. (1959) on an independent sample of patients. In the course of this, the possibility was explored that the signs are a function of brain disorder in general rather than of epilepsy specifically, and the relationships of epilepsy signs to age and intelligence were examined. The latter two questions have not thus far been dealt with in the literature and seem of some possible importance to the interpretation of the signs.

#### METHOD

The sample consisted of 96 patients in a North Carolina Veterans Administration Hospital, including 32 epileptic, 19 other neurological and 45 psychiatric cases. The latter group was made up of 19 psychotics (13 schizophrenics and 6 manic-depressives) and 26 neurotics (comprising anxiety states, hysterics, obsessive-compulsives, and neurotic depressives). All groups bore discharge diagnoses. Diagnoses of epilepsy were based on abnormal EEG tracings and clinically authenticated seizures; other neurological cases included a variety of brain involve-

<sup>1</sup> This investigation was begun while the author was a staff psychologist at the Veterans Administration Hospital, Durham, North Carolina. Thanks are due to Noble David and Russell F. Tomlinson for making the appropriate records available for study. Thanks are also due to Theodore X. Barber and Donald M. Isaac of the Medfield State Hospital for valuable suggestions and critical reading of the manuscript.

ments but had no history of seizures. The psychiatric groups had been inpatients on the psychiatric service of the hospital. The psychiatric groups were constituted by drawing every third relevant case on file in which an MMPI had been given during the hospitalization covered by the final diagnosis. The epileptic and other neurological groups were made up of all patients who met the criteria for inclusion.

Patients' MMPI profiles were scored for the signs given by Hovey et al. (1959). Results were tabulated as percentages of subjects in each group reaching or surpassing cutting points from one to all 14 signs. In this, the psychiatric group was considered both as a whole and in its two component groups. For each between-group comparison, fourfold contingency tables were prepared to show the percentages of patients in the two groups obtaining and not obtaining the maximally discriminating number of signs. These contingencies were evaluated statistically by the chi square test for independent groups.

The correlations of number of signs with age and intelligence were computed by the Spearman rank-order method. Intelligence here is Wechsler-Bellevue verbal IQ, as estimated from Ohio Literacy Test scores according to the regression equation reported by Strong (1959).

#### RESULTS

Table 1 presents the percentages of patients in each group reaching or surpassing all possible cutting points. Validation of the signs requires at some cutting point a relatively high percentage of epileptics and relatively low percentages of patients in other groups. This situation occurs only in the comparison between epileptics and psychotics at the 5- and 6-sign cutting points, where 56%—26% and 44%—16% splits are found. As the table shows, the plurality of epileptics over psychotics at these two points is statistically significant beyond the .05 level of

TABLE 1  
PERCENTAGES OF PATIENTS IN EACH GROUP REACHING OR PASSING VARIOUS CUTTING POINTS

Groups	Number of signs													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Epileptic ( <i>N</i> = 32)	94	88	81	75	56 <sub>b</sub>	44 <sub>b</sub>	31	25	16	3	3	3	0	0
Neurological ( <i>N</i> = 19)	95	90	79	68	53	37	16	10	5	0	0	0	0	0
Psychiatric ( <i>N</i> = 45)	91	82	71	67	44	36	22	22	11	4	4	2	2	0
Psychotic ( <i>N</i> = 19)	89	79	63	63	26 <sub>a</sub>	16 <sub>a</sub>	5 <sub>a</sub>	5 <sub>a</sub>	5	0	0	0	0	0
Neurotic ( <i>N</i> = 26)	92	85	76	69	58	50 <sub>b</sub>	35 <sub>b</sub>	35 <sub>b</sub>	15	8	8	4	4	0

Note.—Entries in the same column with subscripts a and b differ beyond the .05 confidence level.

confidence (minimum  $\chi^2 = 2.84$ ,  $p < .05$  for a one-tailed test.<sup>2</sup>)

Neither the neurotic nor the neurological group is significantly lower in signs than epileptics; for these comparisons the lowest  $p$  value is above .25. In fact, the proportion of neurotics at several of the cutting points is actually higher than the corresponding proportion of epileptics.

Table 1 also shows that the signs statistically differentiate the psychotics from the neurotics at the 6-, 7-, and 8-sign cutting points.

Table 2 gives the correlations of age and estimated verbal IQ with number of signs obtained within each group and for all groups combined. No correlation coefficient reaches an acceptable confidence level. The highest degree of correlation,  $-.25$  among the epileptics between estimated verbal IQ and number of signs, reaches merely the .18 confidence level, for a two-tailed test ( $t = 1.40$ , using the formula presented by Siegel, 1956).

DISCUSSION

The present results indicate that the epilepsy signs proposed by Hovey, Kooi, and Thomas do not perform their intended function in a population of North Carolina veterans. The only relevant discrimination is between epileptics and psychotics, and there are two limitations on the value of this finding.

First, the discrimination seems more associated with uniqueness of the psychotic group than of the epileptic group. Table 1 shows that the psychotics drop off markedly at the 5-sign cutoff point and remain below all other groups up to the 9-sign point. Of these differences

only some with the neurotics are large enough for statistical acceptance, but the average of the psychotic groups' maximum separation from comparison groups is nearly twice that of the epileptic group. Furthermore, the optimal cutting points for differentiating epileptics are less uniform than are those for psychotics. The signs, then, seem somewhat better as negative indices of psychosis than as positive indices of epilepsy.

The second limitation to the discrimination of epileptics from psychotics lies in its relatively low true-positive and high false-positive rates. These are considerably less favorable than those found by the original investigators. Generally speaking, the successful identification of around half the epileptics does not seem worth the price of incorrectly naming around 20% of psychotics as epileptic. Since psychotics are apt to outnumber epileptics in many settings and since the majority of patients diagnosed psychotic are not epileptic, the actual number of false-positive cases in any unselected population would seem likely to dwarf the roughly 50% of epileptics correctly named. The resulting increase in fruitless referrals for epilepsy workup would overload available facilities in most settings to an extent that would make use of the signs impractical.

TABLE 2  
RANK-ORDER CORRELATIONS OF AGE AND ESTIMATED VERBAL IQ WITH NUMBER OF SIGNS OBTAINED

Groups	Age	VIQ
Epileptic ( <i>N</i> = 32)	$-.05$	$-.25$
Psychotic ( <i>N</i> = 19)	$.10$	$.20$
Neurotic ( <i>N</i> = 26)	$-.03$	$-.03$
Other		
Neurological ( <i>N</i> = 19)	$.11$	$.12$
All groups	$.03$	$.01$

<sup>2</sup> The one-sided statistical test was used here and wherever epileptics were compared with other groups, since any tendency of the epileptics to have fewer signs than a control group would be prima facie evidence of the signs' invalidity. All comparisons not involving epileptics were evaluated by two-tailed statistical tests.



As the selection of patients before testing became more accurate, the practical embarrassment of this false-positive rate would decrease. But very high accuracy in weeding out nonepileptics would be required before the actual number of false positives became a negligible percentage of patients judged by the signs. Yet it should be remembered that the signs are intended to assist physician's judgments, not vice versa. Also, little more than half the true epileptics would be identified unless a lower cutting point were chosen, and the present study does not indicate reliable differentiation of epileptics at any lower cutting point.

The crux of the difficulty in using the signs with the present population is the fact that the North Carolina epileptics obtain many fewer signs than Hovey, et al.'s epileptics. The percentage curve of North Carolina epileptics over all numbers of signs is significantly lower ( $\chi^2 = 8.31$ ,  $p < .05$ ) than that of Hovey and associates' (1959) two epileptic groups, by the two-tailed Kolmogoroff-Smirnoff test (Siegel, 1956, p. 135). This is seen more clearly in the fact that Hovey and associates' two epileptic groups had means of 9.6 signs (original group) and 8.1 signs (validation groups), while the present epileptic sample has a mean of 5.2 signs. Comparisons between psychotic and neurotic groups in the two studies show negligible and nonsignificant differences.

Thus it is only in the epileptic samples that the present and the earlier findings are not comparable. The question thus arises of possible differences between the samples on one or more characteristics influencing the tendency to obtain epilepsy signs. The two samples seem similar in some respects: they are both groups of veterans who have required hospitalization, and their epilepsy diagnoses were apparently based on essentially the same criteria. On the variables of age and verbal IQ investigated here, comparable data is not reported for the sample of Hovey et al. (1959), but the low insignificant correlation between each of these variables and number of signs among the present epileptics suggests that neither variable is critical to the difference in signs between the samples.

The two epileptic samples do, however, come from widely separated geographical areas of the United States. Hovey et al. (1959) carried out their study on Utah veterans, while this study was done on North Carolina residents. This distinction may imply a number of differences between the samples, some of which might be relevant to their difference in signs. For example, Negroes constituted 19% of the North Carolina sample,

and are no doubt less strongly represented in the Utah sample. Another possible factor lies in the difference in dominant religious influence in the two areas, Mormon as opposed to Baptist.

Whatever demographic variables may be operating here, it is suggested that their influence might well lie in a difference in attitudes toward epilepsy reflecting itself in different patterns of MMPI responses by epileptics in the two samples. Other patient groups studied here do not seem differentially influenced despite subcultural differences.

The failure to find more general validity to the signs in the present population makes equivocal the finding that organics without seizures are not different in signs from epileptics. In this population it is just one aspect of the signs' inadequacy. However, it leaves open the possibility in the findings of Hovey et al. (1959), that the signs discriminate epileptics because they have brain dysfunction.

The epilepsy signs originated as a means of differentiating true epileptics from patients with "spells," i.e. epileptic-like behavior not accompanied by abnormal brain-wave patterns or clear cut seizures. This study casts no light on the signs' value for this purpose. However, neither does the validation procedure of Hovey et al. (1959). These authors validated their signs on several psychiatric groups<sup>3</sup> not selected for "spells." It is precisely these latter differentiations that the present study cannot confirm. In the original study the epilepsy signs also discriminated patients with paroxysmal and non-paroxysmal EEGs. The present findings also do not bear on these results.

It does seem clear that the epilepsy signs are not universally valid and useful. Where they are valid, the possibility remains that they are signs of organicity, not epilepsy specifically. It may tentatively be suggested that age and intelligence will not prove critical variables in future study.

<sup>3</sup> Hovey et al. (1959) used a third psychiatric group of "personality problems," which has no counterpart in the present study.

## REFERENCES

- HOVEY, H. B., KOOR, K. A., & THOMAS, M. H. MMPI profiles of epileptics. *J. consult. Psychol.*, 1959, 23, 155-159.
- SIEGEL, SIDNEY. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- STRONG, P. N. Correlation between the Ohio Literacy Test, grade achieved in school, and Wechsler-Bellevue IQ. *J. clin. Psychol.*, 1959, 15, 71-72.

(Received July 27, 1962)



## SEX AND PERSONALITY DIFFERENCES IN RELATION TO FANTASY<sup>1</sup>

C. W. LAGRONE

*Texas Christian University*

In an attempt to determine (a) the nature of the daydreams of representative college students, and (b) how variations in these daydreams are related to differences in sex and personality, typical daydreams written in narrative form were collected from 96 women and 123 men. Results show that daydreams of men differ significantly from those of women both with respect to need drive and content area. Evidence that some personality variables among men and women are related to differences in daydreams is suggestive but not conclusive.

Fantasy as obtained from clinical study or via projective methods as applied to normal or abnormal subjects has received a good deal of attention. Despite the data which have been accumulated, little is known, as Singer and Schonbar (1961) have pointed out, concerning the daydreams of normal people and how variations in their daydreams relate to specific personality characteristics. The present study was thus undertaken to gather data relative to two questions: What kinds of daydreams do typical college students have, and how are variations in these daydreams related to differences in sex and personality?

The approach here employed is unique in two ways. First, each subject wrote in narrative form his most typical daydream. Singer and Schonbar used the questionnaire method to study the relation of personality characteristics to reported day and night dreams in women graduate students. Shaffer and Shoben (1956, p. 206) report data concerning daydreams of normal people but they too used the questionnaire method. The second way in which the present study is unique is in the classification of daydreams on two dimensions: need drive and content area. The Shaffer and Shoben data list only types of daydreams, and while some of the types such as "Money or possessions" and "Sexual" clearly refer to content, other types such as "Worry" and "Homage" cannot be identified as to content, nor can the motivation be inferred.

### PROCEDURE

The subjects for the study were 96 women and 123 men who were taking the sophomore level course in General Psychology at Texas Christian University.<sup>2</sup>

<sup>1</sup> This study was supported by the Hogg Foundation, The University of Texas.

<sup>2</sup> Among other reasons, 45 students taking a junior elective course in Abnormal Psychology were excluded from this sample because the means for the

The data were collected anonymously and were obtained from 96% of the members of the classes in the following manner: each student was given an envelope which contained the Minnesota Multiphasic Personality Inventory with answer sheet and two additional sheets of paper. At the top of the first sheet were the following statements:

- (1) Are you male or female? (2) Write in story form your most typical (characteristic, representative, or frequent) daydream or fantasy. Please give as detailed an account as you possibly can.

At the top of the second sheet were the following statements:

In whatever manner seems most appropriate to you, write down as many statements as you can concerning your daydreams. I want to know such things as how often you daydream, the different kinds that you have, how you feel about your daydreams, and everything else about your daydreams that you can remember.

The students were told to complete the materials out of class and to return them on a certain date.

An identifying number was written on each of the different sets of materials. The fantasies were coded independently by two raters. The coding sheet was a double entry table involving need drive on one axis and content area on the other. Thus, each rater checked for each daydream the cell which in his opinion corresponded to the primary need drive and the primary content area in which the need drive was expressed. If a secondary need drive or content area seemed evident, it was also checked. Included also on the coding sheet were three 5-point scales for ratings of Acceptance, Frequency, and Variability of the person's daydreams as indicated from the comments on the second sheet. The raters' coded judgments and the MMPI scores were punched on IBM cards for analysis on a 1620 computer.

women in Abnormal on all of the scales of the MMPI (except the *Mf* Scale) were higher than the means for the women in General Psychology ( $p < .01$ ). Cf. Wise (1959).

## RESULTS AND DISCUSSION

*Rater Agreement*

Agreement between the two raters was fairly high. The correlation for Acceptance was .60, for Frequency .79, and for Variability .64. In 42% of the daydreams, the raters were in agreement concerning both the primary content area and the primary need drive. There was 65% agreement concerning primary need drive and 62% agreement concerning primary content area. Although a secondary need drive and/or content-area was recorded only in about one-fourth of the daydreams by either rater, one rater coded as secondary what the other coded as primary in 5% of the daydreams, both with respect to need drive and content area.

*Sex Differences*

Chi square tests for each rater show that the daydreams of men differ significantly from those of women both with respect to need drive and content area ( $p < .001$  for each rater). With respect to the variables of Acceptance, Frequency, and Variability, only one statistically significant difference was found. Men are more accepting of their daydreams than are women ( $t$  test,  $p < .02$ ).

The manner in which men and women differ in their daydreams can be seen in Tables 1 and 2. In these tables the two raters' codings have been combined so that the totals which are given are twice as large as they would be if each rater were considered separately. As can be seen from Table 1, the most frequent need drive in men is New Experience. For men having New Experience as their primary need drive, the two most frequent content areas in which this need drive is expressed are Adventure or Travel and Sexual Intercourse. The need drive in women which is

the most frequent by a wide margin is Affiliation. For women having Affiliation as their primary need drive, the most frequent content area in which this need drive is expressed is Marriage or Family.<sup>3</sup>

Although large differences between men and women exist in frequencies of specific need drives (for example 50.6% for Affiliation in women as compared to 10.5% for Affiliation in men), the relative frequencies of the various need drives are fairly similar in men and women. The rank-difference correlation between the relative rankings of need drives in men and women is .71. This correspondence, however, does not hold for content areas where the correlation is only .11. The marked difference in rank for content areas between men and women is particularly evident in the case of Marriage or Family, Adventure or Travel, and Money or Possessions.

There are several interesting similarities and differences in the data reported by Shaffer and Shoben and that of the present study. Both studies find Money or Possessions to have a high frequency, but sex differences are much greater in the present study. In the Shaffer and Shoben data Physical Appearance ranks fairly high, especially in women, whereas in the present study it has the lowest rank for both men and women; also sexual daydreams have a much lower frequency in the present study. These variations could have resulted from differences in method, population, or the manner of classifying the daydreams.

Chi square tests for differences in need drive and content area for the raters were made between groups of men and women who were rated high and low on the variables of Acceptance, Frequency, and Variability; those rated as 1 or 2 were compared with those rated as 4 or 5. In these comparisons the following two statistically significant relationships were noted either as the result of combining chi square tests for the two rater's judgments or as the result of the chi square tests being independently significant and in the same direction.

In men there is a difference in content-area between those who are high and low on Variability ( $p < .01$ ). Men who are low in Variability daydream more in the content-areas of Adventure or Travel, Money or Possessions, and Sexual Intercourse than do men high in Variability. Men high in Acceptance daydream more than those low

TABLE 1  
RELATIVE FREQUENCIES AND RANK ORDER OF NEED  
DRIVES IN MEN AND WOMEN

Need drives	Men		Women	
	Total	Rank	Total	Rank
Abasement	5	10	1	11
Achievement	28	4	20	3
Affiliation	37	3	105	1
Aggression	8	8.5	0	12
Autonomy	13	6	2	9.5
Dominance	8	8.5	2	9.5
New Experience	66	1	13	4
Nurturance	2	11.5	7	6
Recognition	50	2	27	2
Seclusion	12	7	4	7
Security	15	5	8	5
Succorance	2	11.5	3	8

<sup>3</sup> The writer will send to the reader upon request mimeographed protocols of representative daydreams and mimeographed tables showing the interrelationship between need drives and content areas among men and women.

TABLE 2  
RELATIVE FREQUENCIES AND RANK ORDER OF CONTENT AREAS IN MEN AND WOMEN

Content areas	Men		Women	
	Total	Rank	Total	Rank
Adventure or travel	37	2	3	10
Death, illness, or serious accident	2	10	10	6
Marriage or family	11	9	81	1
Money or possessions	40	1	8	7
Physical appearance	1	11	1	11
Pre- or extra-marital romance	27	5	28	2
Recreation or hobbies	13	8	4	8.5
Sexual intercourse	24	6.5	4	8.5
Varied content determined by life situations	24	6.5	13	5
Varied content secondary to need drive	36	3	17	4
Vocation	31	4	20	3

in Acceptance in the content-areas of Adventure or Travel, Pre- or Extra-Marital Romance, and Sexual Intercourse; those low in Acceptance daydream more frequently in the area of Money or Possessions ( $p < .01$ ).

#### Personality Differences

In order to determine the manner in which need drives and content areas are related to differences on the various scales of the MMPI, the following comparisons were made. For each of the scales, and for both the men and women considered separately, those individuals whose scores were at least one half a standard deviation below the mean were compared with those whose scores were at least one half standard deviation above the mean. Relatively few of these comparisons were statistically significant, and in view of the large number of comparisons which were made, it may well be that all of the differences are a function of chance.

The chi square tests which were significant ( $p < .05$ ) were three related to need drive and one related to content area in the case of men and two related to content area in the case of women. On the Depression Scale, the men who scored high had more frequently the need drives for Autonomy and Security, whereas among those who scored low, the need drive for New Experience was more frequent. On the Hypomania

Scale, the need drive for Autonomy was more frequent among the men who scored high; whereas the need drive for New Experience was more frequent among those who scored low. On the Social Scale, Recognition was more frequent among men in the high group, whereas New Experience was more frequent in the low group. On a similar comparison for the psychotic profile (*Sc* and *Pa*) the content-area of Money or Possessions was more frequent among the men in the low group ( $p < .01$ ).

For the women the two scales producing statistically significant differences with respect to content-area were the Depression Scale and the Interest (*Mf*) Scale. When the women who scored high on the *D* Scale and on the *Mf* Scale were compared with those who scored low, the content area of Marriage or Family was more frequent among those who scored low on each of these scales.

#### REFERENCES

- SHAFFER, L. F., & SHOEN, E. J. *The psychology of adjustment*. New York: Houghton Mifflin, 1956.  
SINGER, J. L., & SCHONBAR, R. A. Correlates of daydreaming, *J. consult. Psychol.*, 1961, 25, 1-6.  
WISE, L. M. Abnormal psychology as a selective factor: A confirmation and extension, *J. educ. Psychol.*, 1959, 50, 192-194.

(Received August 6, 1962)



## RELATIONSHIP BETWEEN TIME IMAGERY AND RORSCHACH HUMAN MOVEMENT RESPONSES

RONALD B. KURZ

*Syracuse University*

In order to test the hypothesis that M is related to the slow-fast dimension of the time concept, 46 female undergraduates were given the Harrower Multiple-Choice Rorschach Test and the Time Metaphor Test in a group administration. Correlations between M and the ratings of preferences for the metaphors as satisfying descriptions of time revealed that the Ss who produced relatively few M preferred swift time images while Ss who perceived larger numbers of M found the slow or static time images more satisfying. The results were considered support for the notion that M, as a measure of capacity for delayed need satisfaction, is inversely related to the level of motivation for the rapid passage of time.

The perception of moving human beings (M) on the Rorschach has been related to impulse control and the ability to delay the immediate gratification of impulses (Rapaport, Gill, & Schafer, 1946, p. 214). A number of studies have attempted to clarify this relationship and have demonstrated that subjects who produce relatively large numbers of M tend to have greater control over overt motor activity (Singer & Herman, 1954; Singer, Wilensky, & McCraven, 1956) and expression of affective responses (Meltzoff & Litwin, 1956). Apparently the person who has more M in his Rorschach record has greater capacity than the person who has few M to inhibit expression of his impulses and to delay the satisfaction of his needs.

The concept of delayed gratification has also been related to time concept or time sense. Siegman (1961) has shown a relationship between impulse control, time estimation and orientation, and delinquent behavior. Siegman found that a group of delinquents had a shorter future time perspective and lower time estimation score than a group of nondelinquents, and that among the delinquents there was a positive correlation between the subject's future time perspectives and their scores on a task of motor impulse control. Le Shan's (1952) finding that lower-class children had shorter future time perspectives than middle-class children and the findings of Barndt and Johnson (1955) and of Davids, Kidder, & Reich (1962) that delinquent children have shorter future time perspectives than nondelinquents may be considered further

evidence for the relationship between time variables and the ability to delay the immediate satisfaction of impulses.

In the present study an attempt was made to investigate the relationship between the delayed gratification aspect of M and the slow-fast dimension of the time concept. Taking a motivational approach (Siegman, 1961), it may be assumed that the person who has difficulty in delaying the immediate satisfaction of his needs for the sake of some other and more distant goals is highly motivated for time to pass quickly. Waiting is uncomfortable for him; time must pass swiftly so that he may be gratified without excessive delay. Conversely, it may be assumed that the person who is well able to withstand delays has little need for time to pass rapidly. The difference in motivation should be reflected in the ways a person thinks about or conceptualizes time. Specifically, the person who has little capacity for delayed gratification ought to find it more satisfying to think about time in rapid-moving terms than in slow terms while the person with relatively more capacity for delayed gratification should find the slower terms more satisfying. In regard to the Rorschach, one would expect persons who produce relatively few M to prefer to conceptualize time in swifter terms than persons with relatively large numbers of M.

Recent work by Knapp and Garbutt (1958) has resulted in the development of a set of metaphors or images concerning time called the Time Metaphor Test, by means of which the

TABLE 1  
CORRELATIONS OF RATINGS OF THE PREFERENCE FOR EACH METAPHOR WITH M

Metaphor	Correlation with M	Metaphor	Correlation with M
a galloping horseman	+ .48**	a devouring monster	.00
a fleeing thief	+ .40**	budding leaves	-.02
a fast moving shuttle	+ .35*	an old woman spinning	-.04
marching feet	+ .31*	a stairway leading upwards	-.04
a speeding train	+ .30*	a string of beads	-.06
a whirligig	+ .29*	a winding spool	-.07
a bird in flight	+ .24	an old man with a staff	-.11
a dashing waterfall	+ .19	a quiet, motionless ocean	-.23
a tedious song	+ .03	a road leading over a hill	-.29*
a space ship in flight	+ .01	a burning candle	-.30*
a vast expanse of sky	.00	wind-driven sand	-.33*
a large revolving wheel	.00	drifting clouds	-.34*
a massive glacier	.00		

\*  $p < .05$ .

\*\*  $p < .01$ .

subjects may indicate their most satisfying and least satisfying ways of conceptualizing time. This technique was used in the present study to measure the degree to which the subjects preferred slow or rapid images of time. It was predicted that the subjects with relatively few M would find swift images more satisfying as ways of conceptualizing time, while the subjects who perceive relatively large numbers of M would find the slower images more satisfying.

#### METHOD

In a single group testing session, 46 female undergraduate psychology students were given the Harrower Multiple-Choice Rorschach Test (Harrower & Steiner, 1945), and the Time Metaphor Test (Knapp & Garbutt, 1958).

The Rorschach Test was administered by projecting slides of the 10 blots on a screen in a semi-darkened room. Each subject had an answer booklet which listed 30 possible responses for each blot. The 30 choices for each blot were divided into three groups of 10 choices each. The instructions required the subjects to study the blot, read through each of the three groups of answers for the blot, then to underline one answer in each group which they thought was the best description of the inkblot or any of its parts. The subjects were required, therefore, to underline three answers for every blot. Each blot was exposed for 3 minutes. The Rorschach data were scored for M in the manner prescribed by Beck, Beck, Levitt, and Molish (1961). That is, only choices which clearly indicated human action were considered M choices. Human content

alone was not considered M. For example, on Card II "Two people playing pat-a-cake" was scored M, but "Two clowns" was not.

The Time Metaphor Test was administered next. This test consisted of 25 images (see Table 1) which might be considered poetic descriptions of time (Knapp & Garbutt, 1958). The subjects were presented with the list of 25 metaphors, preceded by the following instructions:

Below are listed 25 phrases which might be employed by a poet or a writer to symbolize his sense of time. I should like you to read through this list of phrases and then indicate before each how appropriate you think this phrase is in evoking for you a satisfying image of time. First select the five phrases that seem to you most satisfying and before each place the number "1". Then pick out the next five most satisfying phrases and before them place the number "2". Continue this process until you have placed the number "5" before the five least satisfying phrases in your opinion.

#### RESULTS

The mean number of M responses for the group was 4.0, with a standard deviation of 1.65.

In order to test the hypothesis that M is related to the swiftness of the time images, product-moment correlations for the rating of each metaphor with M score were computed. These correlations are given in Table 1 in ranked order. It can be seen that there is a marked tendency for images involving swift movement to be correlated positively with M,

while static images and those involving slow movement are correlated negatively with M. To clarify possible confusion about the positive and negative correlations in the table, it should be pointed out that because of the procedures used by the subjects in responding to the metaphors, the least preferred metaphors have the highest numerical values. A positive correlation therefore indicates that high M people placed the swift metaphors lowest in rank order. In other words, the subjects who gave relatively more M on the Rorschach preferred to describe time by using static or slow-moving images while the subjects with relatively few M preferred the images involving rapid movement. It should be noted that all of the top 10 metaphors involve motion and that perhaps only 2 of them, "marching feet" and "a tedious song," do not express swiftness. Among the last 10, possibly only "an old woman spinning" and "a winding spool" express rapid motion, while the rest are either static or slow. Chi square with Yates correction, comparing swift images with slow or static images among the 10 most positively and 10 most negatively correlated metaphors, was 5.0,  $p < .05$ .

As a result of a factor analysis of the 25 metaphors, Knapp and Garbutt (1958) described three distinct clusters. In the present study the relationships between these clusters and M were investigated. The first cluster, termed by Knapp and Garbutt the Dynamic-Hasty Cluster, was characterized by images of rapid movement. This cluster included the following metaphors: a galloping horseman, a fleeing thief, a fast moving shuttle, a speeding train, a whirligig, a dashing waterfall, a space ship in flight. When median ratings by each subject for this cluster of metaphors were correlated with M, the  $r$  was  $+ .52$ ,  $p < .001$ .

The second cluster, called the Naturalistic-Passive Cluster, was characterized by references to nature and lack of rapid motion, and consisted of the following metaphors: drifting clouds, wind-driven sand, a road leading over a hill, a quiet, motionless ocean, budding leaves, a vast expanse of sky. The product-moment correlation between M and the median ratings on the second cluster was  $- .41$ ,  $p < .01$ .

The third cluster, named the Humanistic Cluster, seemed to involve human or human surrogate figures or human artifacts, and included the following: a tedious song, a large revolving wheel, a devouring monster, an old woman spinning, a string of beads, a winding

spool, an old man with a staff, a burning candle. Median ratings on the Humanistic Cluster correlated  $.00$  with M.

## DISCUSSION

The data of the present study seem to support the hypothesis that the subjects who report seeing relatively more human movement in ink blots prefer to conceptualize time in slow or static terms, while those who report seeing relatively little human movement prefer to conceptualize time in terms of swift movement. These data appear to be related to the findings that M is related to the ability to delay impulse gratification (Meltzoff & Litwin, 1956; Singer & Herman, 1954; Singer & Spohn, 1954; Singer et al., 1956). The high M person is capable of delaying satisfaction of his needs for relatively long periods of time without undue discomfort. For him, the rapid passage of time is unimportant in the eventual gratification of his needs and he finds it satisfying to think of time as stable or moving slowly ahead. However, the orientation of the low M person is one of quick sequences of tension and relief. He must achieve gratification of his impulses within a brief period of time and he prefers to conceptualize time as passing rapidly so that he can be released from discomfort quickly.

An alternate explanation of the data which does not take account of the slow-fast dimension could be made by referring to the "poetic" quality of the metaphors. It could be maintained that the swift images are less poetic or imaginative than the slow or static ones. Since the subjects who produce numerous M are said to be more creative or imaginative (Rorschach, 1942), they would naturally prefer the slow or static metaphors. However, this explanation does not seem to be supported by the finding of zero correlation between M and the images of the Humanistic Cluster which appear to be as poetic as those in the Naturalistic-Passive Cluster.

The lack of relationship between M and the Humanistic Cluster is itself a surprising finding. It would be expected that the subjects who show sufficient interest in human activity to perceive it in meaningless blots would also show interest in describing time in similar terms. However, the data suggest that the slow-fast dimension of time is probably the most important source of common variance between M and the metaphors and that interest in human activity is unimportant. Further investigation of the relationship



between M and other time variables, such as time estimation and future time perspective, is now being carried out.

The exclusive use of women subjects in this study raises the question of the generality of the results. Certainly if men had been available for and used in the study, limits in the ability to generalize to both sexes would not be a problem. Although the impression of the writer (and apparently that of many other husbands who have waited for their wives) is that the sexes do differ in their attitudes and motives concerning time, the absence of empirical findings regarding this point leaves the question of generality an unanswered one.

### REFERENCES

- BARNDT, R. J., & JOHNSON, D. M. Time orientation in delinquents. *J. abnorm. soc. Psychol.*, 1955, 51, 343-345.
- BECK, S. J., BECK, A. G., LEVITT, E. E., & MOLISH, H. B. *Rorschach's test: Vol. 1. Basic processes*. New York: Grune, 1961.
- DAVIDS, A., KIDDER, C., & REICH, M. Time orientation in male and female juvenile delinquents. *J. abnorm. soc. Psychol.*, 1962, 64, 239-240.
- HARROWER, M. R., & STEINER, M. F. *Large scale Rorschach techniques*. Springfield, Ill.: Charles C Thomas, 1945.
- KNAPP, R. H., & GARBUTT, J. T. Time imagery and the achievement motive. *J. Pers.*, 1958, 26, 426-434.
- LESHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, 47, 582-589.
- MELTZOFF, J., & LITWIN, D. Affective control and Rorschach human movement responses. *J. consult. Psychol.*, 1956, 20, 463-465.
- RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing: The theory, statistical evaluation, and diagnostic application of a battery of tests*. Vol. 2. Chicago: Year Book Publishers, 1946.
- RORSCHACH, H. *Psychodiagnostics*. Bern: Hans Huber, 1942.
- SIEGMAN, A. W. The relationship between future time perspective, time estimation, and impulse control in a group of young offenders and in a control group. *J. consult. Psychol.*, 1961, 25, 470-475.
- SINGER, J. L., & HERMAN, J. Motor and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, 18, 325-331.
- SINGER, J. L., & SPOHN, H. Some behavioral correlates of Rorschach's experience type. *J. consult. Psychol.*, 1954, 18, 1-9.
- SINGER, J. L., WILENSKY, H., & MCCRAVEN, V. G. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, 20, 375-383.

(Received September 4, 1962)

## BRIEF REPORTS

### CHARACTERISTICS OF HOMOSEXUALLY INVOLVED INCARCERATED FEMALES<sup>1</sup>

WILLIAM G. MILLER AND THOMAS E. HANNUM

*Iowa State University*

The purpose of this study was to examine differences between a group of homosexually involved and a group of nonhomosexually involved female prisoners.

A prisoner was included in the homosexual group only if it were certain that she was actively engaged in homosexual relationships while incarcerated and in the heterosexual group only if it were certain that she had avoided any homosexual relationships during incarceration. Due to lack of any heterosexual choice, the degree of true homosexuality could not be determined.

The data were collected from testing, case history data, and counseling interviews. The interviews and group selections were made by the psychologist regularly employed at the institution.

The data included scores from the Minnesota Multiphasic Personality Inventory, Kuder Preference Record, Wide Range Achievement Test, and Otis Quick-Scoring Mental Ability Test, all routinely administered upon admission. In addition, questions pertaining to homosexuality were asked and many items of a biographical nature were taken from the case history.

A factor analysis using intercorrelations between people on variables was pursued. An attempt was made to cluster the intercorrelations rationally into homosexual and heterosexual groups. This analyses failed to provide any evidence of factors common to either group and

was not pursued further. Rather the means of each group in each variable was computed for further description and analysis.

Comparison of the groups on the variables indicated a high degree of similarity in almost all comparisons. The MMPI mean *T* scores, although generally high, were not significantly different. The MMPI profiles, coded by Welsh's method, provided interesting but nondifferentiating information.

A chi square test of the degree to which the groups differed in the distribution of occurrence of the 10 MMPI scales did indicate a significant difference existed in the case of Scale 1, the Hypochondriasis scale, the heterosexual groups scoring higher.

The comparisons of Kuder scores and profiles indicated greatest disparities were on the Social Service scale with the heterosexuals scoring higher and the Music Scale with the homosexuals scoring higher.

Examination of the remainder of the variables furnished little to differentiate the groups except for two direct questions concerned with the involvement in and approval of homosexual relations. The fact that the homosexually involved persons stated their approval does not, of course, aid us in diagnosing homosexuality in persons reluctant to provide such information.

Though none of the variables incorporated in this study differentiated the homosexual group from the heterosexual group to a degree sufficient for individual diagnosis, it is possible that other psychological measurement devices could be found to be more discriminative. It is also quite possible, or even probable, that homosexuality is such a complex characteristic that no one such device will be found. A combination of several less discriminative measures may have to suffice.

(Received July 16, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Thomas E. Hannum, Department of Psychology, Iowa State University, Ames, Iowa, or for a fee from the American Documentation Institute. Order Document No. 7398 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit \$2.50 for microfilm or \$1.75 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

## FANTASIZED DANGER AS A FUNCTION OF PARENT-CHILD CONTROLLING PRACTICES<sup>1</sup>

DONALD W. TIFFANY AND FRANKLIN C. SHONTZ

University of Kansas

Rogers (1959) proposed a "continuum which reaches from rigidity and fixity of psychological functioning on the one hand, to psychological flow and changingness on the other" (p. 96). Previous work by the authors reported an attempt to operationalize this continuum through the measurement of *experienced control* (C) in pre-adolescents via the Picture Q Technique.

In the present study, degree of C served as the criterion for subject grouping. High C scores represented the restricted and rigid end of Rogers' psychological continuum, while low C scores represented the fluid, free, and comfortable end. It was hypothesized that children who experience high C (rigidity and fixity) reveal a characteristic mode of response to the environment, viz., they evidence a basic distrust and fear of unfamiliarity.

The subjects were 22 preadolescent boys randomly selected from a Midwestern grade school. They were administered the Picture Q Technique (PQT) and a fantasy task called Mother Hubbard. The PQT consists of 48 drawn parent-child interactions. There are two pictures of each specific parent<sup>2</sup>-child interaction, in one the parent is male and in the other, female. The pictures range from extreme permissiveness of parent-child interaction to parental control of the child, such as inhibiting the child's behavior. The child is asked to describe himself by Q sorting the pictures on a 5-point scale. The weight assigned to each picture is determined by its placement. Total C scores are computed for the

24 C pictures by summing the weights assigned to these pictures.

For the Mother Hubbard task the child was given 5 minutes to think of and write down the consequences of Mother Hubbard finding her cupboard bare. The responses were typed on 4 × 6 inch cards, shuffled, and given to three clinical psychologists (judges) with the instructions to rank order the cards on the basis of the degree of *physical danger* expressed in each response.

The 11 children with high C scores (median 58; range 56-65) formed one group while the 11 children with low C scores (median 54; range 54-55) formed the other. The groups were closely comparable in age, intelligence, and school grade.

Kendall's coefficient of concordance for large samples ( $d/ = 21$ ) for the judge's ratings of the Mother Hubbard responses was significant ( $W = .67$ ;  $p < .01$ ).

The Mann-Whitney  $U$  test was used to test the independence of the two groups on the basis of the weights assigned to the Mother Hubbard responses by the judges. It was found that the two groups differed significantly ( $U = 29.5$ ;  $p < .05$ ; two-tailed test). This finding supports the hypothesis that children whose self-sort reveals high parental C fantasize more environmental danger than children whose self-sort reveals low parental C. Confirmation of the hypothesis lends support to Rogers' contention that degree of permissiveness in parent-child interaction is a significant determinant in the child's perceptions of himself and his environment.

### REFERENCE

- ROGERS, C. R. A tentative scale for the measurement of process in psychotherapy. In E. A. Rubinstein, and M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959. Pp. 96-107.

(Received August 6, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Donald W. Tiffany (Psychology Department, University of Kansas, Lawrence, Kansas) or for a fee from the American Documentation Institute. Order Document No. 7458 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> The term "parent" also refers to parent surrogates.



## CROSS-VALIDATION OF STEIN'S SYMBOL-GESTALT TEST FOR BRAIN DAMAGE<sup>1</sup>

OSCAR A. PARSONS,<sup>2</sup> HARRIET I. MASLOW

*University of Oklahoma School of Medicine*

AND KENNETH D. STEWART

*University of Nebraska*

Stein (1961) recently reported highly successful discrimination of brain damaged subjects from non-brain-damaged subjects by use of a rapid, easily administered, age-corrected, perceptual-motor task, the "Symbol-Gestalt Test." While the test has been used by Stein on a number of samples in prepublication studies, questions may be raised concerning the generalizability of Stein's findings to different populations, examiners, and locales.

In the course of a series of studies of brain damage, it was possible to test the applicability of the Symbol-Gestalt Test to two populations drawn from widely separated Veterans Administration hospitals, Durham, North Carolina and Oklahoma City, Oklahoma.

The population for the Durham study was comprised of 20 clearly diagnosed cortically brain damaged patients (BD) and 20 non-brain-damaged patients (NBD); the Oklahoma City study consisted of 20 BD patients and 30 NBD patients. The diagnoses of patients in both BD groups fell mainly in the CVA, tumor, or traumatic head injury categories. The NBD had diagnoses predominantly of neurosis or psychophysiological reaction. All patients were male hospitalized veterans with a mean age of 41.3 years and a mean education level of 9.7 years; the groups were equated on these variables.

The Symbol-Gestalt Test followed the administration of the Vocabulary subtest of the WAIS.

<sup>1</sup> An extended report of this study may be obtained without charge from O. A. Parsons, University of Oklahoma Medical Center, 800 N.E. 13th Street, Oklahoma City 4, Oklahoma, or for a fee from the American Documentation Institute. Order Document No. 7459 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

This research was supported, in part, by PHS Research Grants B-1459 and B-2507 to the senior author from the National Institutes of Neurological Diseases and Blindness, USPHS.

<sup>2</sup> Oscar A. Parsons and Harriet I. Maslow are also affiliated with the Oklahoma City Veterans Administration Hospital.

The test protocols were scored according to instructions appearing in the manual; inter-scoring reliability was 95%.

The results in both studies indicate a clear-cut difference between the performance of the BD and NBD. In the Durham population the BD attained a mean score of  $-.668$  and the NBD a mean score of  $-.042$  ( $t = 2.56$ ;  $p < .02$ ). In the Oklahoma City study the BD attained a mean score of  $-.482$  and the NBD a mean score of  $-.059$  ( $t = 4.23$ ;  $p < .001$ ). Stein utilized a cutoff point of zero to differentiate BD subjects from NBD subjects: positive scores (+) fall in the NBD category and negative scores (-) in the BD category. It should be noted the four mean scores in the present study carry a negative sign. Therefore, the application of Stein's cutoff point to these populations results in a high percentage of false positives (31%).

In attempting to account for these results, it is important to note that while the groups used in our study were in many ways comparable to Stein's, e.g., Veterans Administration patients, sex, age, diagnoses, etc., their educational level was lower by approximately 1.5 years and IQ by 14 points. In the Durham BD group, Stein performance is not correlated with either education or vocabulary. However, in the three other groups the correlations are positive and significant (four of six correlations are above  $+.69$ ). Correlations of this magnitude suggest that education and IQ play a greater role than Stein has emphasized.

These findings indicate that the Symbol-Gestalt Test is sensitive to impairment associated with cortical brain damage. However, standardization of the test on populations which have a greater range of education and intelligence is indicated before it can be used as a clinical tool with diagnostic power.

### REFERENCE

- STEIN, K. B. The effect of brain damage upon speed, accuracy and improvement in visual motor functioning. *J. consult. Psychol.*, 1961, 25, 171-177.

(Received August 15, 1962)

## EFFECT OF INTELLECTUAL LEVEL AND NEUROPSYCHIATRIC STATUS ON THE DIVERSITY OF INTENSITY OF SEMANTIC DIFFERENTIAL RATINGS<sup>1</sup>

CHARLES NEURINGER

*University of North Dakota*

The hypothesis that a person's capacity to perceive concepts in a varied manner (as measured by the diversity of semantic differential ratings) is related to intellectual level was evaluated by Ware (1958) with disappointing results.

Since previous research seemed to indicate that utilization of intensity rating spaces appears to be related to the emotional state of the individual, it was felt that further exploration of the relationships between intellectual level, neuropsychiatric disposition and diversity of semantic differential intensity rating spaces would prove to be fruitful.

An 18 concept, 9 scale semantic differential, and the Information Subtest of the Wechsler-Bellevue Intelligence Scale, Form I were administered to neuropsychiatric hospital patients ( $N=16$ ) and to normal hospitalized patients ( $N=16$ ). The subjects were matched within +1 or -1 correct responses on the Information Subtest in order to control for intellectual level when examining the effects of neuropsychiatric status on intensity diversity of semantic differential ratings. The extreme, moderate, mild and neutral rating spaces on each scale were assigned a value of 3, 2, 1, and 0, respectively. The mean standard deviation of the intensity ratings were computed for each subject from the 162 scale judgments made by that subject. The greater the value of the mean standard deviation, the greater the amount of intensity diversity of ratings on the semantic differential.

A  $t$  test of the mean standard deviation differences between the normal and the neuro-

psychiatric groups (1.16 and .85, respectively) yielded a value of 2.844 (significant at the .009 level). A Spearman Rho between the mean standard deviations and Information Subtest scores of the 32 subjects was found to be .07 (significant at the .72 level). The subjects were also divided into two groups via a median split of the Information Subtest scores. The mean standard deviations for the above and below median groups were 1.04 and .97, respectively. A  $t$  test of the differences yielded a value of .583 (significant at the .55 level). The same procedure was followed with the upper and lower quartiles on the Information Subtest. The mean standard deviations for the upper and lower quartile subjects were 1.03 and .98, respectively. A  $t$  test yielded a value of .312 which was significant at the .76 level of confidence. It would appear that diversity of intensity of semantic differential ratings seems to be related more to neuropsychiatric involvement than to intellectual level.

Although one would expect that intelligence should be related to the ability to perceive and produce scores reflecting diversity of intensity, this does not seem to be the case in this study. It is strongly felt here that failure to find such a relationship may be due to the error of considering intelligence as a unified entity and thus accepting single intelligence scores as reflecting that unity. The effects of emotional disturbance on perception of intensity diversity may be related to the neurotic's need to organize his world in a simple manner, making it more manageable. In this sense, lack of intensity diversification may be related to the same dynamics that produce rigidity in perception and thinking.

### REFERENCE

- WARE, E. E. Relationships of intelligence and sex to diversity of individual semantic meaning spaces. Unpublished doctoral dissertation, University of Illinois, 1958.

(Received June 22, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Charles Neuringer (Psychology Department, University of North Dakota; Grand Forks, North Dakota) or for a fee from the American Documentation Institute. Order Document No. 7456 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.



## DIGIT SPAN PERFORMANCE AS A FUNCTION OF NOXIOUS STIMULATION<sup>1</sup>

SANDRA PYKE AND NEIL McK. AGNEW

*University of Saskatchewan*

Easterbrook (1959) proposed that increased drive would reduce range of cue utilization. Such a reduction will facilitate performance on simple tasks (those involving a small number of cues) and impair performance on complex tasks, leading to a drive by task complexity interaction.

Seventy-five female university students were assigned at random to three equal groups. All subjects were given two testings, on alternate forms, of the digits forward and digits backward tests. Group I was given a first testing under normal conditions with no shock or threat of shock but, just prior to the second testing, electrodes were placed on the fingers of the non-preferred hand and a 100-volt, 30-millisecond shock was administered. Group II was shocked, then given a first testing followed by removal of the electrodes and a second testing with no shock or threat of shock. When performing under shock condition, in addition to receiving a sample shock, experimental subjects were informed that poor performance would lead to further shocks. Group III served as a control group to provide an estimate of improvement due to practice.

The Taylor Manifest Anxiety Scale was administered to all subjects. To monitor the duration of the effects of the shock and nonshock conditions, continuous heart rate recordings were taken from an additional 17 subjects not part of the original sample.

*Scoring.* (a) The number of digits in the series preceding the one in which the individual made his first mistake was the number correct score. (b) The sum of the number of series incorrectly repeated for each testing was the error score.

<sup>1</sup> An extended report of this study may be obtained without charge from Neil McK. Agnew, Psychological Research Center, 1214 College Drive, Saskatoon, Saskatchewan, or for a fee from the American Documentation Institute. Order Document No. 7460 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Group I (nonshock, shock) and Group II (shock, nonshock) provide data suitable for a switchback analysis (Brandt, 1938) which tests for combined stress effects on first and second testing against combined nonstress effects on first and second testing. Only in the case of digits forward number correct was a significant high stress impairment obtained ( $t = 2.98$ ,  $p < .01$ ,  $df 48$ ). Using an analysis of covariance procedure, there was no evidence of stress effects on second testing on any score. In a  $2 \times 2$  analysis of variance high anxiety subjects show significant impairment in both stress and nonstress conditions only on the digits forward data, number correct, and only on first testing.

Using a Friedman two-way analysis of variance, it was found that shock, plus threat of shock, significantly increased heart rate during performance ( $p < .001$ ,  $k = 4$ ,  $n = 10$ ), regardless of whether the stress condition was introduced on first or second testing.

The limited effects of anxiety, whether defined in terms of noxious stimulation or questionnaire scores, raise some questions regarding the clinical application of the digit span tests.

The results obtained suggest a drive by practice interaction since, when shock was introduced after some practice on the task, no consistent evidence of a drive or an anxiety effect appeared. Assuming that a practiced task is a less complex one, the findings are in keeping with the proposed drive by task complexity interaction. Increased drive may impair performance only when a task is complex or is relatively unpracticed.

### REFERENCES

- BRANDT, A. E. Tests of significance in reversal or switchback trials. *Res. Bull., Ia. Agric. Exp. Sta.*, 1938, No. 234, 60-97.  
EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, 66, 183-201.

(Received September 17, 1962)



## FURTHER SUPPORT OF SOME FINDINGS ABOUT THE CHARACTERISTICS OF SMOKERS AND NONSMOKERS

BRUCE C. STRAITS  
University of Chicago

AND

LEE SECHREST  
Northwestern University

Smoking and nonsmoking are alternative and opposite behaviors of sufficient complexity and potential social importance to be of increasing interest to social scientists. A recent review of research findings relating to smoking behavior pointed succinctly to the psychological aspects of the problem as a fruitful area for research endeavor (Matarazzo & Saslow, 1960). It is the purpose of this paper to report some data which tend to support previously reported findings.

Two questionnaires developed for use in the study consisted of a 270-item true-false "personality inventory" comprised of items from a number of well-known personality scales plus items from a scale used to measure the extent to which an individual feels that what happens to him is dependent upon his own behaviors as opposed to the operations of chance (internal versus external control). The second questionnaire was of a factual, biographical nature. The subjects for the study were 125 male college students initially, plus an additional 120 to check a near-significant finding concerning internal-external control. Of the original 125 subjects, 54 identified themselves as smokers.

The major findings may be summarized as follows:

1. Smokers were significantly higher on the MMPI *Pd* scale and significantly lower on the *Sr* scale.
2. Smokers were more "chance oriented" ( $p < .08$ ) in the first sample, a finding which was consistent at the .10 level in the second sample.
3. Consistently with previous research but not at a significant level smokers were found to have lower MMPI *L* and *Si* scores.
4. Smoking was significantly related to both

coffee and beer drinking as previously reported. However, the *Pd* and *Sr* scales were related only to beer, not to coffee drinking. Smoking was not related, as might be expected from an oral indulgence hypothesis, to consumption of milk or soft drinks.

5. From the standpoint of biographical variables smokers, when compared with nonsmokers, were found to: attend church less often; come more often from private preparatory schools; be more frequent among Catholics; be more likely to belong to social fraternities. One finding which was expected but not confirmed was that smokers were found to be more frequent in families in which both parents were reported to smoke.

The finding that smokers are more "chance oriented" than nonsmokers is particularly intriguing inasmuch as it offers some hint concerning the manner in which heavy smoking may be maintained in the face of strong evidence of its danger. In effect the smoker may reduce the *dissonance* between his smoking and medical reports by adhering to the philosophy that all life is a gamble.

In general the findings are supportive of previous descriptions of the typical smoker as a convivial person lacking strong identification with cultural standards, and of the nonsmoker as less at ease socially, sober, and imbued with a strong ethic.

### REFERENCE

- MATARAZZO, J. D., & SASLOW, G. Psychological and related characteristics of smokers and nonsmokers. *Psychol. Bull.*, 1960, 57, 493-513.

(Received September 24, 1962)

# JOURNAL OF CONSULTING PSYCHOLOGY

August 1963

Vol. 27, No. 4

## Contents

### ARTICLES

- Photos Preference Test (PPT) as a Measure of Mental Status for Hospitalized Psychiatric Patients: *Bernard Meer and Albert H. Amon* - - - - - 283
- Field Dependence and Overcoming Embeddedness: *Stephen A. Karp* - - - - - 294
- Comparative Efficiency of Some Tests of Cerebral Damage: *Maurice Korman and Stanley Blumberg* - - - - - 303
- Patients and Therapists Assess the Same Psychotherapy: *Herman Feifel and Janet Eells* - - - - - 310
- Body Cathexes of Parents of Normal and Malformed Children for Progeny and Self: *Louise Centers and Richard Centers* - - - - - 319
- Psychological Correlates of Somatic Complaints in Pregnancy and Difficulty in Childbirth: *Marvin Zuckerman, John I. Nurnberger, Sprague H. Gardiner, James M. Vandivoer, Beatrice H. Barrett, and Arie den Broeijen* - - - - - 324
- Respiration and GSR as Functions of White Sound in Schizophrenia: *Vladimir Pishkin and David Her- shiser* - - - - - 330
- Affective Stimuli and Disturbance of Thought Processes: *Gayle Deering* - - - - - 338
- The D 48 Test as a Measure of General Ability among Grade School Children: *Harrison G. Gough and George Domino* - - - - - 344
- Scaling the Scales: Use of Expert Judgment in Improving the Validity of Questionnaire Scales: *K. Warner Schaie* - - - - - 350
- Counselor Anxiety in Relation to Amount of Clinical Experience and Quality of Affect Demonstrated by Clients: *Peter D. Russell and William U. Snyder* - - - - - 358

### NOTES AND COMMENTS

- Childhood and Intercurrent Intellectual Performance of Adult Schizophrenics: *George W. Albee, Ellen A. Lane, Clare Corcoran, and Ann Werneke* - - - - - 364
- Failure to Validate the Cr and Sm Scales of the MMPI: *John L. Vogel* - - - - - 367

### BRIEF REPORTS

- Spontaneous GSR and Anxiety Level in Sociopathic Delinquents: *Robert Fox and Walter S. Lippert* - - - - - 368
- Social Desirability Response Set in the Individual: *Erling E. Boe and William Kogan* - - - - - 369
- TAT and MMPI Psychopath Deviant Scale Differences between Delinquent and Nondelinquent Adolescents: *Albert W. Silver* - - - - - 370
- Development of Potentially Therapeutic Interpersonal Relations in Schizophrenics through Competition: *Albert E. Myers* - - - - - 371
- Kent E-G-Y: Differential Scoring and Correlation with the WAIS: *Gerald L. Clore, Jr.* - - - - - 372
- Experimental Prediction of the Long-Term Effect of Psychiatric Therapy: *Gudmund J. W. Smith and Gunner Johnson* - - - - - 373
- Errata - - - - - 374

# Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

## Specifications

**Brief Report.** The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the

author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

<sup>1</sup> An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Documentation No. — from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$— for microfilm or \$— for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

**Extended report.** Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (a) The extended report should be typed with single spacing for economy in duplication. (b) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (c) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.



## PHOTOS PREFERENCE TEST (PPT) AS A MEASURE OF MENTAL STATUS FOR HOSPITALIZED PSYCHIATRIC PATIENTS<sup>1</sup>

BERNARD MEER

AND

ALBERT H. AMON

*Stockton State Hospital*

*University of North Carolina*

The Photos Preference Test (PPT) consists of 100 facial photographs, standardized for like and dislike. This study attempted to determine whether deviant responses to the photographs were related to psychological disturbance. 490 newly admitted patients to a state hospital were tested on admission and retested 4 months later or before then if discharged sooner. 369 "normal" controls were given the PPT. The results showed: (a) patients give more deviant responses than controls ( $p < .01$ ); (b) schizophrenics give more deviant responses than nonschizophrenics ( $p < .0001$ ); (c) patients giving less deviant responses on admission are discharged sooner than patients giving more deviant responses ( $< .01$ ); and (d) female patients show a decrease in deviant responses as they improve ( $p < .01$ ), while males do not.

Much of the research with human faces as a projective device stems from Murray's (1933) early study on the projection of fear and maliciousness and from the introduction of the Szondi Test in this country by Deri (1949). Although the latter test was initially greeted with a great deal of enthusiasm, recent studies have found little evidence to support the assumptions underlying its use (Borstelmann & Klopfer, 1953; David, Orne, & Rabinowitz, 1953; Gordon, 1953; Gordon & Lambert, 1954; Hurley, 1957; Mussen & Krauss, 1952; Short, 1954). Nevertheless experience with the Szondi Test has encouraged investigators to find new ways of using pictures of human faces as projective tests.

By requesting subjects to describe the personality (Izard, 1952) or emotional state (Glad & Shearn, 1956) of the persons in the photographs, investigators have tried to learn about the personalities of the perceivers. A more frequent model for using facial photographs as a projective test is one that requests subjects to respond first in terms of a like-dislike or favorable-unfavorable dimension and then on a second occasion to attribute personality characteristics (Chambers,

1957a; Flyer, 1951; Liggett, 1957), values (Fensterheim & Tresselt, 1953), and occupational categories (DeRath, 1957) to the same photographs. An analysis is then made for characteristics attributed to the liked and disliked pictures in order to determine to which variables a subject reacts positively and to which he does not. A third model and one most closely related to the present study is that of obtaining normative reactions to the photographs and then measuring the deviations from the norms for different clinical groups (Chambers, 1957b; Eysenck, 1954).

In our work with a series of 100 pictures of human faces, we found marked individual differences in the number of pictures liked when groups of normal subjects were asked to give their first impression as to whether they liked or disliked each of the 100 photographs. Included in our sampling of normals were professional and nonprofessional hospital employees, old people, college students, student nurses, labor groups, PTA groups, church groups, etc. No matter what comparisons were made between these groups, the stimulus value of the photographs remained relatively the same, that is, the rank order of the photographs in terms of a like-dislike dimension changed very little from one group to the next. It is this high degree of consensus about the affective value of the various facial photographs that suggested the present study.

<sup>1</sup> This study was supported by a grant from the Department of Mental Hygiene, State of California, and the first of the two studies reported here was read at the Western Psychological Association meetings in San Jose, California, in April 1960.

It is interesting to note that some years ago Bell (1948, p. 287), in reviewing the literature on expressive movements, suggested the idea of constructing a test based on the social stereotypes that photographs provide. More recently Eysenck (1954) and Chambers (1957b) used this technique to differentiate normals from clinical groups. Eysenck, in fact, pointed out that it was the pictures which produce simple stereotyped responses that appear from a diagnostic point of view to promise more success than difficult non-stereotyped photographs.

The fact that there is a great deal of agreement as to the affective quality of the photographs indicates that the cues evoking these reactions are responded to by a great part of the general population. It is our proposition that the failure to interpret these cues in the usual way may be indicative of the presence of a psychological disturbance. On this basis the following hypotheses were made:

1. That hospitalized psychiatric patients will show more deviation in their affective reactions to the photographs than a "normal" control group.
2. That patients on admission whose affective reactions to the photographs are more deviant from the standardization group will improve less quickly than those patients showing less deviation.
3. That patients who qualify for discharge or indefinite leave of absence within 4 months following admission will show a greater decrease in deviation from the standardization group upon retesting than patients who do not so qualify.

#### *Photos Preference Test (PPT)*

The Photos Preference Test consists of 100 facial photographs which are the full-face views from the booklets prepared by Campbell and Burwen (1956) for another purpose. The photographs include 10 pictures of each sex for four different age groups (adolescent, young adult, middle age, and old age). Another group of photographs of 20 middle-aged men were included to increase the reliability of the test. The test requires that subjects give their first impression as to whether they

like (L) or dislike (D) each of the people in the photographs. The 100 pictures are in booklet form and can be administered individually or to a group.

#### *Standardization Group*

Approximately 500 male and 500 female "normal" adults and adolescents were tested in Stockton, California, a community of about 100,000 people. *Normal* was defined as anyone not institutionalized at the time of testing. Subjects were asked to give their age, education, and occupation but were told to omit their name. All testing was voluntary. A small percentage of subjects, about 5%, refused to take the test. A standardization group consisting of 200 males and 200 females was selected from the above subjects in such a way that the distribution of the standardization group in terms of age (15 years and up) and education is similar to that of the 1950 United States Census. This standardization group was used to determine the stimulus value of each photograph.

To determine the stability of the stimulus values of the photographs, the 200 protocols for each sex were assigned to groups of 100 randomly, but constrained to the same age and education distributions. The stimulus values of the photographs based on each set of 100 protocols were compared. The rank-order correlations were .95 for the two male populations and .96 for the two female populations, confirming our informal findings about the reliability of the stimulus values of the photographs. Because we found consistent sex differences between males and females to certain age-sex categories, we decided to retain separate norms for males and females despite the high overall correlation between them.

#### *Scoring the Photos Preference Test (PPT)*

In this report the chief concern will be with an overall congruence score, "C," which indicates the similarity of a subject's pattern of Ls and Ds to that of the standardization group of his sex.

First, two scale values were obtained for each photograph, using a modification of the method of successive categories (Torgerson,



1958). These values represent the degree to which each photograph was liked by the male and female standardization groups, respectively.

For an analysis of the responses of each individual, a method similar to one reported by McReynolds (1951) was used. This method has the advantage of yielding a C score which is independent of the proportion of Ls and Ds, and which provides for a weighting of each deviation from the consensus, according to its degree of deviation. In brief, assume a subject gives 40 Ds, he will get a score of 100 or perfect congruence if his Ds are assigned to the 40 least-liked photographs. Any deviation from this pattern reduces his congruence score. A weighting procedure for the extent of the deviation is used. We then total all the deviation scores, find what percentage this is of the maximum possible deviation, and subtract from 100 to obtain a final C score.

Since the C score tends to become less reliable the more extreme the proportion of likes and dislikes, subjects who gave fewer than 10 Ls or 10 Ds were not scored for C.

### Reliability

Two measures of reliability were obtained: (a) the split-half reliability coefficient for 96 female patients is .84, and (b) the test-retest reliability coefficient for 103 student nurses, corrected for restricted range, is .83.

### PROCEDURE

This report covers two separate studies. Study I was carried out between November 1958 and October 1959, and Study II between June 1960 and February 1961. Since the latter was an attempt to replicate the former, the procedure used in Study II followed closely that of Study I. The slight differences that did occur will be mentioned later.

### Subjects

*Study I.* For a period of approximately 6 months all new admission patients at Stockton State Hospital who were no more than 50 years old were considered potential subjects for this study. Excluded, however, were patients diagnosed as alcoholics, narcotic addicts, or as having any one of the organic disorders. Of the 134 male and 173 female patients who were available for the study, 70 or 52% of the males and 96 or 55% of the

females met all criteria and gave protocols scorable for C.

Following Study I, a group of approximately 60 male and female new admission patients were given only the Photos Preference Test. The procedure was the same as in Study I except for the fact that this group of patients was told that the investigators were interested in how groups of patients as a whole responded to the test and were not interested in how any specific individual answered his test. Therefore, they were told to omit their name on the answer sheet. Since all normal subjects were given these instructions in contrast to the patients of the first study, it was felt necessary to determine how important was the effect of identifying oneself. The C scores and the number of like and dislike responses of these patients were compared with a group of patients from the first study, equated for age, education, and diagnosis. When no appreciable differences were found, the "nameless patients" were combined with the patients of Study I. These patients were not retested and therefore could not be used to test Hypothesis 3.<sup>2</sup>

*Study II.* Approximately 9 months following the completion of the first study, all new admission patients for a period of 5 months were tested. This time the age limit was raised to 55 years. Otherwise the criteria remained the same. Of the 187 males and 205 females who were available for this study, 122 or 65% of the males and 131 or 64% of the females gave protocols scorable for C.

Table 1 shows the distribution, by diagnosis, of all male and female patients used in the two studies.

Control subjects consisted of males and females drawn from normals who had not been selected for the standardization groups. Within the framework of matching the controls with the patient groups for age and education, the selection was carried out on a random basis. Table 2 shows the mean and standard deviation of age and education for male and female patients and controls.

### Test Procedure for Patients

*Study I.* All patients were tested in relatively small groups (15-25 patients) within 1 week of their admission. The patients were brought to the testing room by the ward personnel. Before the testing began, the patients were told that they were participating in a research study<sup>3</sup> and that the results of the testing would be kept confidential.

<sup>2</sup> There were enough identifying data on the answer sheet (age and occupation) to determine the name of each patient. This the investigators did in order to obtain these patients' diagnoses and to determine whether they were short- or long-stay patients. The investigators have no way of knowing whether any of the patients realized that his protocol could be identified even without his name.

<sup>3</sup> We wish to acknowledge our appreciation to Robert B. Freeman who was responsible for the testing in Study I and for his general assistance in the first half of the project.



TABLE 1  
DISTRIBUTION OF PATIENTS IN STUDIES I AND II  
COMBINED BY DIAGNOSIS<sup>a</sup>

Diagnosis	Male	Female
Schizophrenic		
Paranoid	51	50
Acute undifferentiated	22	30
Chronic undifferentiated	36	35
Schizo-affective	8	37
Other	11	11
Total schizophrenic	128	163
Nonschizophrenic, psychotic		
Involucional	9	30
Manic depressive	2	20
Depressive	6	0
Paranoid state	1	1
Nonschizophrenic, nonpsychotic		
Neurotic	28	34
Personality trait disturbance	27	9
Personality pattern disturbance	12	6
Sociopathic personality disturbance	6	3
Other	2	3
Total nonschizophrenic	93	106

<sup>a</sup> Based on the American Psychiatric Association's Diagnostic and Statistical Manual of 1952.

It was impressed on the patients that the test results would have no influence on the length of stay or the nature of their treatment in the hospital. The purpose of these instructions was to lessen the anxiety of the patients and make the test conditions more equal to that of the control group. Three tests were administered: the Photos Preference Test, the Draw-a-Person Test, and the MMPI, in

that order. The instructions for the PPT were the same as those given to the normals. Some patients protested that they could not judge whether they liked a person or not just by looking at a picture. In such cases they were told, "Write down your first impression," or "Do the best you can." In some cases where a patient was either especially reluctant or appeared to be having difficulty writing, the examiner helped out by filling in the answers as given by the patient. Because a follow-up was to be made of the patients, they were asked to fill in their name on the answer sheet.

Patients who were given a discharge or went on indefinite leave of absence within 4 months of their admission were retested with the PPT and Draw-a-Person Test at the time of their departure. Patients who remained in the hospital for a period of 4 months were retested at that time. The 4-month cutoff was chosen to separate the "short-" and "long-stay" patients because experience had shown that approximately half of the patients left the hospital in the first 4 months of their hospital stay. Patients who left the hospital within 4 months had to remain out of the hospital a minimum of 30 days to be rated short-stay patients.

*Study II.* Except for the fact that there was a change in the test administrator (Amon), the procedure for the first part of Study II followed closely that of Study I. Approximately one-third of the way through the second study, however, it was decided to administer the PPT twice in order to determine the consistency of the patient's responses, and so identify those patients who were responding randomly. It should be noted that only the data for the first administration of the PPT were used in the comparisons to be reported. The tests that were administered from this point on were the PPT, a 40-item exploratory questionnaire, a repeat administration of the PPT, and the MMPI, in that order. The Draw-a-Person Test was dropped because it was time consuming and because the results of the first study indicated that it showed little promise. As in Study I these patients were

TABLE 2  
MEAN AND STANDARD DEVIATION OF AGE AND EDUCATION FOR MALE AND  
FEMALE PATIENTS AND CONTROLS

Subjects	Study I					Study II				
	N	Age		Education		N	Age		Education	
		M	SD	M	SD		M	SD	M	SD
Males										
Patients	99	31.6	9.5	10.7	2.8	122	35.4	11.1	10.3	3.2
Controls	99	32.2	10.2	10.8	2.8	108	34.7	11.4	10.6	2.9
Females										
Patients	138	35.4	9.4	10.7	2.6	131	36.5	10.3	10.5	2.8
Controls	138	36.0	10.1	11.2	2.5	131	36.6	11.1	11.1	2.3

Note.—Some of the low education controls in Study I were used in Study II.

TABLE 3  
MEAN C SCORE DIFFERENCES BETWEEN PATIENTS AND CONTROLS

Studies	Males						Females					
	N	Pa- tients	Con- trol	N	t	p*	N	Pa- tients	Con- trol	N	t	p*
Study I	99	75.7	78.5	99	1.71	<.05	138	79.3	82.0	138	2.58	<.005
Study II	122	75.6	79.5	108	2.78	<.003	131	77.4	81.9	131	4.17	<.0001
Studies I & II combined												
Education <11 years	108	73.9	78.7	57 <sup>a</sup>	2.43	<.008	114	77.3	82.1	61 <sup>b</sup>	2.98	<.002
Education ≥11 years	113	77.3	79.2	110	1.41	<.08	155	79.2	81.8	141	2.82	<.005

\* One-tailed test.

<sup>a</sup> Because an insufficient number of normals were available as controls at the low educational level, two separate analyses were made: (a) for subjects with less than 11 years of education and (b) for subjects with 11 or more years of education.

retested at the time of discharge or leave of absence or at 4 months following admission, whichever came first.

### RESULTS

The results shown in Table 3 clearly support the first hypothesis. In every comparison between patients and normals, equated for age and education, the former have lower C scores, thus showing more deviation in responses to the photographs than the latter. In addition the findings in Table 3 show that the difference in C scores between patients and normals is more striking among females than males and among the less educated than among the more educated groups.

One of the factors operating to reduce the significance of the results in Table 3 was the

great variability of the patients' scores.<sup>4</sup> Some of the highest and lowest C scores were those obtained from patients. We were interested in determining whether these C scores were in fact related to an independent measure of mental status. For this purpose, the patients were divided into those diagnosed as schizophrenic in one group and all other diagnostic categories in a second group. Since schizophrenia is a label given to the most pathological patients, we predicted that their C scores would be significantly lower than for the nonschizophrenic patients. Table 4 gives the

<sup>4</sup> Standard deviations were in all cases significantly greater for patients than for normals. However, these differences were not so large as to call into question the validity of the *t* test (Boneau, 1960).

TABLE 4  
MEAN C SCORE DIFFERENCES BETWEEN SCHIZOPHRENIC AND NONSCHIZOPHRENIC PATIENTS

Studies	Males					Females				
	N	Schizo- phrenic	Non- schizo- phrenic	N	t	p*	N	Schizo- phrenic	Non- schizo- phrenic	p*
Study I	60	73.2	79.4	39	2.45	<.008	81	77.2	82.5	<.0007
Study II	68	72.2	80.0	54	3.89	<.0001	82	75.4	80.7	.002
Studies I & II combined	128	72.7	79.7	93	4.50	<.0001	163	76.3	81.7	<.0001

\* One-tailed test.

TABLE 5

MEAN C SCORE DIFFERENCES AT THE TIME OF ADMISSION BETWEEN SHORT- AND LONG-STAY PATIENTS

Studies	Males						Females					
	N	Short-stay	Long-stay	N	t	p*	N	Short-stay	Long-stay	N	t	p*
Study I	54	78.4	73.1	37	2.14	<.02	83	80.9	76.4	53	2.72	<.004
Study II	80	77.7	73.5	32	1.86	<.04	95	77.8	75.8	33	.96	ns
Studies I & II combined	134	78.0	73.3	69	2.86	<.003	178	79.2	76.2	86	2.39	<.009

Note.—Short-stay patients left the hospital within 4 months of admission while long-stay patients did not leave the hospital within 4 months of admission.

\* One-tailed test.

results of this comparison for each study and for both combined. The difference in C scores between schizophrenic and nonschizophrenic patients is extremely significant in each study and for both sexes. As predicted, schizophrenic patients as a group show more deviation (lower C score) than the nonschizophrenic patients.

A further analysis of the patient data revealed that patients diagnosed as schizophrenic tend to be less educated than patients not so diagnosed. In order to determine how much education contributes to the differences in C scores between schizophrenics and nonschizophrenics, correlation coefficients between education and C score were calculated for all patients. These correlation coefficients are positive and significant, that is, the more educated patients tend to have higher C scores. However, the degree of relationship between these two variables is so slight,  $+ .125$  ( $p < .05$ ) for males and  $+ .189$  ( $p < .01$ ) for females, that it accounts for very little of the variance in C scores between schizophrenic and nonschizophrenic patients.

Finally, it should be pointed out that while normals have significantly higher C scores than patients (Table 3), their C scores are no different from those of the nonschizophrenic patients (Table 4). In short, it is the deviant responses of the schizophrenic patients that make for the difference between normals and patients.

The second hypothesis predicted that patients giving less deviant responses to the facial photographs on admission would leave the hospital sooner than patients giving more

deviant responses. Results relevant to this hypothesis are shown in Table 5. Focusing on the individual studies separately, we find that three out of the four  $t$  tests are significant, and the fourth (Study II, Females) while not significant is in the predicted direction. The  $t$  tests for the combined studies are significant at the .003 level for males and at the .009 level for females. Thus patients who get high C scores on the PPT when tested during the first week of their admission to the hospital tend to leave the hospital sooner than patients who get low C scores.

Since the C score is significantly related to length of stay and also to diagnosis<sup>5</sup> (see Table 4), the question was posed whether diagnosis is also related to length of stay, and if it is, whether the C score length of stay relationship obtains within the schizophrenic and nonschizophrenic groups. The answer to the first question is shown in Table 6. The results show that diagnosis is indeed related to length of stay, more strikingly for males ( $p < .001$ ) than for females ( $p < .05$ ). In general, patients who do not receive a schizophrenic diagnosis have a much better chance of leaving the hospital within 4 months of admission than patients receiving a schizophrenic diagnosis.

Table 7 shows the relationship between C scores and length of stay<sup>6</sup> within each

<sup>5</sup> In this study the term "diagnosis" refers to the differentiation of schizophrenic from nonschizophrenic patients.

<sup>6</sup> Length of stay for this analysis was the number of days in the hospital up to a year before leave of absence or discharge.



TABLE 6  
RELATIONSHIP BETWEEN DIAGNOSIS AND SHORT-STAY AND LONG-STAY PATIENTS

Studies	Males				Females			
	Short-stay	Long-stay	$\chi^2$	$p$	Short-stay	Long-stay	$\chi^2$	$p$
Study I								
Schizophrenic	36	25	10.11	<.01	57	24	1.07	ns
Nonschizophrenic	44	7			37	10		
Study II								
Schizophrenic	25	27	8.32	<.01	44	37	4.70	<.05
Nonschizophrenic	29	8			40	15		
Studies I & II combined								
Schizophrenic	61	52	18.69	<.001	101	61	4.89	<.05
Nonschizophrenic	73	15			77	25		

Note.—See footnote 5.

diagnostic group and for both groups combined. First it should be noted that all 18 correlation coefficients in Table 7 are negative, that is, the higher the C score the shorter the length of stay. Of particular importance in this table are the correlations within the diagnostic groups for Studies I and II combined. The results show that the correlation between C and length of stay within schizophrenics, although quite modest, is significant

for males ( $r = -.219$ ,  $p < .02$ ) and females ( $r = -.153$ ,  $p < .05$ ). Within the nonschizophrenic group this relationship is significant for females ( $r = -.189$ ,  $p < .05$ ) but not for males, although the correlation is in the predicted direction and close to significance ( $r = -.156$ ,  $p < .10$ ).

The third hypothesis predicted that patients who improve sufficiently to qualify for leave of absence or discharge within 4 months

TABLE 7  
RELATIONSHIP BETWEEN C SCORES AND LENGTH OF STAY FOR DIFFERENT DIAGNOSTIC GROUPS

Studies	Males				Females			
	N	r	t	$p^*$	N	r	t	$p^*$
Study I								
Schizophrenic	54	-.309	2.25	<.05	81	-.202	1.82	<.05
Nonschizophrenic	36	-.259	1.55	ns	55	-.258	1.97	<.05
All patients	90	-.345	3.28	<.02	136	-.256	2.98	<.002
Study II								
Schizophrenic	57	-.150	1.12	ns	80	-.133	1.18	ns
Nonschizophrenic	50	-.065	0.45	ns	46	-.109	0.72	ns
All patients	107	-.181	1.87	<.05	126	-.137	1.54	ns
Studies I & II combined								
Schizophrenic	111	-.219	2.30	<.02	161	-.153	1.94	<.05
Nonschizophrenic	86	-.156	1.43	ns	101	-.189	1.89	<.05
All patients	197	-.268	3.75	<.0001	262	-.187	3.03	<.002

\* One-tailed test.

TABLE 8  
DIFFERENCES IN CHANGE OF C SCORES FOR SHORT- AND LONG-STAY PATIENTS

Studies	Males				Females			
	N <sup>a</sup>	Mean change	F	p	N <sup>a</sup>	Mean change	F	p <sup>a</sup>
Study I								
Short-stay	12	-1.31	.55	ns	32	+3.1	8.32	<.005
Long-stay	22	+1.62			25	-3.6		
Study II								
Short-stay	39	+1.90	1.33	ns	47	+4.90	1.88	ns
Long-stay	26	-0.53			27	+2.60		
Studies I and II combined								
Committed patients only								
Short-stay	29	+1.34	.32	ns	50	+6.94	23.55	<.0001
Long-stay	40	+0.28			39	-0.29		
Committed & voluntary patients								
Short-stay	51	+0.84	.00	ns	79	+4.18	10.36	<.001
Long-stay	48	+0.78			52	-0.39		

<sup>a</sup> Retesting of short-stay patients depended on our being notified by the wards that patients were leaving the hospital. Quite often the ward personnel were too involved with more immediate problems than to think of contacting us. Also, quite a few patients left the hospital on weekends or after hours. For the above reasons, approximately 50% of the short-stay patients were not retested. This explains our reduced N.

\* One-tailed test.

of their date of admission (short-stay patients) when retested at the time of their departure will show a greater decrease in deviation from the standardization group than will patients not so improved (long-stay patients). The latter were retested 4 months following their admission. Table 8 shows the results of an analysis of covariance of the change in C scores between short-stay and long-stay patients, controlling for the effect of the original C score. Here a striking sex difference occurs. For male patients there appears to be no relationship whatsoever between a change in C score and length of stay. On the other hand, this relationship is very significant for female patients. It should be noted that among females this relationship is more significant for committed patients ( $F = 23.55$ ,  $p < .0001$ ) than for committed and voluntary patients combined ( $F = 10.36$ ,  $p < .001$ ). This is understandable since voluntary patients can leave the hospital without necessarily being judged as improved, while committed patients have no such freedom.

### Other Findings

Since the MMPI was administered only once in both studies, that is, at the time of admission, its interest lay in its relationship to length of stay and to the C score. Several analyses were carried out to determine whether the MMPI is predictive of length of stay. The short-stay and long-stay patients were compared for each of the 12 standard scales. Of the 48  $t$  tests (Studies I and II, males and females) only 5 were significant at the .05 level, and no one scale consistently differentiated the two groups in both studies. In those instances where a scale showed a significant difference in one study, the means of that scale were either reversed or showed no difference in the other study. Another approach was to have five experienced clinicians evaluate the MMPI profiles and rank order them in terms of "degree of pathology." This was carried out for Study I only. Since the clinicians showed a high degree of consistency in their rankings, each MMPI profile was assigned the average

rank as its score. These average ranks were then correlated with length of stay, and the results yielded low positive correlations which, however, were not significant. Finally an item analysis for short-stay and long-stay patients was carried out for both studies. A number of items differentiated the two groups of patients for Study I. However, practically none of the items that were significant in the first study showed up as significant in the second study. The failure to find any consistency in the two studies was true for the male and the female patients and discouraged any further use of other scales to predict length of stay. The overall results suggest that whatever the MMPI measures at the time of admission, it bears little relationship to length of stay, at least for this population. An item analysis of the MMPI for patients with high and low C scores similarly yielded negative results.

The Draw-a-Person Test which consisted of two drawings, one male and one female, was ranked by six clinicians<sup>7</sup> for degree of pathology. The data for the male and female patients were treated separately. Since the rankings were fairly reliable, they were combined to obtain an average rank for each set of drawings. These average rank scores were then compared with C and length of stay. The results showed no significant relationship between the Draw-a-Person Test and the other two variables for males or females.

### DISCUSSION

The results of this study support, for the most part, the general hypothesis that psychological disturbance is reflected in deviant affective responses to pictures of human faces. The hospitalized mental patients do indeed show more idiosyncratic responses to the photographs than do the controls. However, it is the more seriously disturbed patients, the schizophrenics, that account for this difference. The results, therefore, indicate that psychological disturbance has to reach very grave proportions before the ability to

respond to facial photographs is markedly affected. In order to determine whether education, too, was related to deviant responses to facial photographs, correlations were carried out between C and education for schizophrenics and nonschizophrenics, separately and for the controls. Only the male and female schizophrenic correlation coefficients were significant ( $+ .19$ ,  $p < .05$  for the males and  $+ .17$ ,  $p < .05$  for the females). The correlation coefficients for the nonschizophrenics were  $+ .05$  for the males and  $+ .08$  for the females, and for the controls  $+ .11$  for the males and  $- .04$  for the females. One may conclude from the above that the ability to react to stimuli such as facial photographs in a way that is shared by most people is a relatively stable characteristic and is upset only under great psychological stress (schizophrenia). In addition, the data suggest that schizophrenics with a limited education may have a little more difficulty resisting perceptual disorganization than schizophrenics with more education.

The second hypothesis in this study is really a question of the prognostic value of the C score, without regard to the type of treatment given to the patients. In a recent review of the literature on the prognostic use of psychological tests, Fulkerson and Barry (1961) point out that where tests are successful in predicting outcome it is because they accurately measure the variable "severity of illness." The results in this study are certainly in line with this generalization. The C score has proved itself as a sensitive discriminator of the more from the less severely ill at the time of admission and has been able to predict length of stay to a moderate but significant degree despite the shortcomings of length of stay as a criterion of improvement.<sup>8</sup> The results have also shown that, although diagnosis and C both correlate with length of stay, C makes some

<sup>8</sup> Length of stay, which has been used as a criterion of improvement in many studies (Fulkerson & Barry, 1961), while objective, is far from being an ideal criterion because some patients stay or leave the hospital for reasons other than the state of their psychiatric condition. A significant variable that has been little studied but which plays an important role in determining length of stay is "family resources."

<sup>7</sup> We wish to acknowledge our appreciation to Charles E. Dutton, Francis McFarland, Rollin Rose, Irving Roy, Emanuel Skoor, and Richard Wolton for their aid in ranking the MMPIs and DAPs.



contribution to the prediction of length of stay independent of diagnosis.

Was the C score effective in differentiating the short-stay from the long-stay patients because the sicker patients were responding autistically or because, not being able to attend to the task, they were responding randomly? To answer this question, most of the patients in the second study were given the test twice in the space of 1 hour. When the number of changes in response to the particular photographs was more numerous than could occur by chance once in a hundred times, the patient was judged inconsistent. The proportion of patients who were found to be inconsistent was surprisingly low, 10% of the males and 11% of the females. The critical question was whether the proportion of short- and long-stay patients was different for the inconsistent and consistent responders. The chi square turned out insignificant ( $\chi^2 = .54$ ), which indicated that the C score's sensitivity to differentiating short- and long-stay patients is *not* based on the fact that long-stay patients are simply more random in their responses to the photographs than short-stay patients.

The results of the third hypothesis, assuming that the difference in findings for the males and females is reliable (Table 8), are most puzzling. Why should the change in C score reflect the improvement of the short-stay female patients but not the short-stay male patients? We tried to rationalize these results by suggesting that since men tend to resist psychiatric treatment more than women and since there is more economic pressure on men than on women from the environment, men are released from the hospital as soon as they manifest a superficial improvement in behavior. Women, on the other hand, may be kept longer until a more significant change in their behavior takes place. If the above were true, we should expect that males would be released more quickly than females (controlling for diagnosis) and that the recidivism rate would be greater for the former than the latter. We found neither expectation to be confirmed in our sample.

A rationalization that is more speculative than the above but which may have more

merit in explaining our results concerns the difference in emphasis that men and women place on the interpersonal role. Early in life boys begin to direct their interests and are directed by society toward things mechanical and scientific, toward ideas and abstractions, in contrast to girls whose interpersonal interests remain central. This difference is not only reflected in the choice of toys, games, and school subjects, but also in the degree of interpersonal communications about feelings and people, as shown by Jourard and Lasakow (1958). Support for this view comes from our standardization data which show that females tend to agree with each other much more than men do as to which facial photographs they like and which they dislike. In short, they share their feelings about likes and dislikes of people to a greater extent than men do. Thus, because of the centrality of the interpersonal needs of women, their ability to respond to the photographs appropriately is likely to fluctuate with the degree of disorganization of central processes. On the other hand, because of the more varied involvements of men, the kind of interpersonal affect reflected in response to the photographs is less central, and therefore their response to the facial photographs would be less likely to fluctuate under central disorganization. The above would explain why male patients even when they improve show little change in C score while women show a striking change. These speculations are suggestive of some important differences between the interpersonal needs of males and females that might reward study.

#### REFERENCES

- BELL, J. E. *Projective techniques: A dynamic approach to the study of the personality*. New York: Longmans, Green, 1948.
- BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, 57, 49-64.
- BORSTELMANN, L. J., & KLOFFER, W. G. The Szondi test: A review and critical evaluation. *Psychol. Bull.*, 1953, 50, 112-132.
- CAMPBELL, D. T., & BURWEN, L. S. Trait judgments from photographs as a projective device. *J. clin. Psychol.*, 1956, 12, 215-221.
- CHAMBERS, J. L. Identification with photographs of people. *J. consult. Psychol.*, 1957, 21, 232-234. (a)

- CHAMBERS, J. L. Trait judgment of photographs by neuropsychiatric patients. *J. clin. Psychol.*, 1957, 13, 393-396. (b)
- DAVID, H. P., ORNE, M., & RABINOWITZ, W. Qualitative and quantitative Szondi diagnosis. *J. proj. Tech.*, 1953, 17, 75-78.
- DERATH, G. Personalized rating responses: An indirect method for obtaining affective reaction to various occupational groups. Unpublished master's thesis, Trinity University, 1957.
- DERI, SUSAN K. *Introduction to the Szondi test*. New York: Grune & Stratton, 1949.
- EYSENCK, H. J. Probleme der diagnostischen Untersuchung und Demonstration des Charakter-Interpretationstestes. *Z. exp. angew. Psychol.*, 1954, 2, 1-32.
- FENSTERHEIM, H., & TRESSELT, MARGARET E. The influence of value systems on the perception of people. *J. abnorm. soc. Psychol.*, 1953, 48, 93-98.
- FLYER, E. S. The development of the picture choice test, an objective projective personality test. Unpublished doctoral dissertation, University of Colorado, 1951.
- FULKERSON, S. C., & BARRY, J. R. Methodology and research on the prognostic use of psychological tests. *Psychol. Bull.*, 1961, 58, 177-204.
- GLAD, D. D., & SHEARN, C. R. An emotional projection test. *Percept. mot. Skills*, 1956, 6, 1-12. (Monogr. Suppl. No. 1.)
- GORDON, L. V. A factor analysis of the 48 Szondi pictures. *J. Psychol.*, 1953, 36, 387-392.
- GORDON, L. V., & LAMBERT, E. J. The internal consistency of the Szondi "factors." *J. soc. Psychol.*, 1954, 40, 67-71.
- HURLEY, J. R. Psychodiagnostic limitations of Szondi interseries changes. *J. clin. Psychol.*, 1957, 13, 396-399.
- IZARD, C. E. Perceptual responses of paranoid schizophrenic and normal subjects to photographs of human faces. Unpublished doctoral dissertation, Syracuse University, 1952.
- JOURARD, S. M., & LASAKOW, P. Some factors in self-disclosure. *J. abnorm. soc. Psychol.*, 1958, 56, 91-98.
- LIGGETT, J. A non-verbal approach to the phenomenal self. *J. Psychol.*, 1957, 43, 225-237.
- McREYNOLDS, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, 46, 131-141.
- MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, 4, 310-329.
- MUSSEN, P. H., & KRAUSS, SHIRLEY R. An investigation of the diagnostic validity of the Szondi test. *J. abnorm. soc. Psychol.*, 1952, 47, 399-405.
- SHORT, P. L. Experiments on the rationale of the Szondi test. *J. ment. Sci.*, 1954, 100, 384-392.
- TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received August 29, 1962)

## FIELD DEPENDENCE AND OVERCOMING EMBEDDEDNESS<sup>1</sup>

STEPHEN A. KARP

*State University of New York, Downstate Medical Center*

Witkin and his colleagues have suggested that field dependence involves ability to overcome the effects of embedding contexts. The present study evaluates this hypothesis in relation to an alternative one, that field dependence involves ability to resist distraction. 150 male college students were given a battery of 18 tests including tests of field dependence, WAIS subtests, Adaptive Flexibility measures, and measures of ability to resist distraction. A factor analysis was carried out on the matrix of intercorrelations between these variables. The results tend to support the Witkin hypothesis; the field dependence and other embeddedness tests load and define different factors than the distraction tests. The two kinds of factors do, however, tend to be moderately correlated.

Since 1948, Witkin and his associates have carried out an extensive series of studies of perceptual field dependence (see Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954, for a summary of the earlier work). On the basis of evidence obtained from these studies, they describe a broad dimension of psychological functioning which they call *differentiation* (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). They demonstrate that from knowledge of one's relative level of differentiation, predictions can be made regarding aspects of his functioning in a variety of areas including perception, problem solving, intelligence, and personality. Among the measures employed to tap the differentiation dimension, a series of tests of perceptual field dependence have been most frequently and centrally used by these investigators. In view of this central role assigned the field dependence dimension in their work, Witkin et al. have carried out a number of studies for the purpose of defining the kinds of cognitive abilities measured by tests of field dependence.

On the basis of evidence obtained from some of their early studies of field depend-

ence, Witkin and his colleagues hypothesized a general cognitive style by which performance in many perceptual and intellectual situations might be characterized. They viewed this cognitive style as reflecting ability to find or manipulate an item which is embedded within a surrounding context. To test this hypothesis, several studies were carried out for the purpose of determining relationships between measures of field dependence and other kinds of measures which, upon rational analysis, appeared to require ability to overcome embeddedness. Witkin and his colleagues predicted that such measures should be significantly correlated with and factorially similar to measures of field dependence.

In one such study, Goodenough and Karp (1961) gave the Wechsler Intelligence Scale for Children (WISC) and three tests of field dependence, the rod-and-frame test (RFT), body-adjustment test (BAT), and embedded-figures test (EFT) to two groups of children (aged 10 and 12). On the basis of rational analysis of the WISC subtests and previous factor analytic studies, they identified three subtests (Block Design, Picture Completion, and Object Assembly) as involving ability to overcome embeddedness. They predicted that these subtests, along with the tests of field dependence, would define a single factor on the basis of the common requirement of overcoming embeddedness. This was substantiated by the results of two factor analyses (one for each group). Two other predicted factors emerged

<sup>1</sup>Based upon the author's dissertation at New York University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is indebted to Jacob Cohen, sponsor of the dissertation, and to Richard Loeff for his assistance in gathering and analyzing the data. This study was supported in part by a research grant (M-628) from the National Institutes of Health, United States Public Health Service, with H. A. Witkin as principal investigator.



in both studies, a Verbal Comprehension factor (loading Vocabulary, Information, Similarities, Arithmetic, and Comprehension) and a Memory or Attention-Concentration factor (loading Digit Span, Arithmetic, and Coding), neither of which consistently and substantially loaded the embeddedness tests.

In another study, Goodman<sup>2</sup> postulated relationships between measures of field dependence and Flexibility of Closure (Botzum, 1951; Pemberton, 1952; Thurstone, 1944, 1949), based upon her evaluation of measures of Flexibility of Closure as involving ability to overcome embeddedness. She obtained significant correlations between the two types of measures. Gardner, Jackson, and Messick (1960) have since found tests of field dependence and Flexibility of Closure to define a single factor.

In a third study, Witkin et al. (1962) predicted that tests of field dependence would be significantly related to tests of Adaptive Flexibility (Guilford, Frick, Christensen, & Merrifield, 1957). This was based upon their view that the latter measures all involved problem situations where the "given" provided an embedding context from which some piece of information had to be extracted and restructured in order to solve the problem. They found two such measures of Adaptive Flexibility, Match Problems and Insight Problems, to correlate significantly with EFT, BAT, and RFT.

The evidence from these studies suggests considerable interrelation and overlap between measures of field dependence, Flexibility of Closure, and Adaptive Flexibility. Witkin et al. have suggested, as a basis for this overlap, that ability to overcome the effects of embedding contexts is a common requirement of tests defining each of these factors. In contrast, interpretations of Flexibility of Closure by Thurstone (1944), Botzum (1951), and Pemberton (1952) suggest that the tests loading this factor require the subject to overcome the effects of distracting, rather than specifically embedding, contexts. The latter interpretation might imply that an embedding context is but a special case of a distracting context.

<sup>2</sup> Unpublished study, described in Witkin et al. (1962).

Evidence from two studies might be considered as supportive of the Thurstone interpretation. Podell and Phillips (1959) and Jackson (1955) each obtained significant correlations between the EFT and a measure which appeared to reflect ability to overcome the effects of distracting, rather than embedding, contexts.

The study of Podell and Phillips involved cluster analysis of a group of cognitive tests, including the EFT and Podell's (1957) Form Embeddedness Test. The latter test requires the subject to locate a series of simple figures in each of two kinds of complex figures, one organized and one random. Their results suggest that the EFT and a total time measure of Form Embeddedness (for both random and embedding contexts) correlate highly with one cluster (Spatial Decontextualization). A second Form Embeddedness measure (ratio of time taken to find figures in embedding contexts to time taken to find them in random contexts) correlates little with this cluster. This would imply that EFT performance may be as highly related to performance under conditions of distraction as it is to performance under conditions of embeddedness.

Jackson administered a battery of tests to small groups of male and female subjects. The battery included the EFT and a Words in Noise Test. In the latter situation, subjects were asked to identify words hidden by background white noise. This background of noise can be viewed as providing a distracting context. Jackson obtained a significant relationship between the tests for men ( $r = .46$ ,  $p < .05$ ) but not for women ( $r = .21$ ). Although a different perceptual modality was used in this study—and results were obtained for men but not for women—these findings offer some support for viewing embedding and distracting contexts as related.

The above results suggest that further evidence is needed before the issue of distraction versus embeddedness can be settled. The present study is designed to provide such evidence. It must be noted, however, that this study is not directed at investigation of the broad issues of how or why one may be distracted. Rather, the kinds of distraction situations chosen for study all

involve irrelevant visual stimuli which surround or intersect critical test items which must be located or manipulated by the subject. The distraction situations are thus similar to embeddedness situations, such as EFT, where critical visual items are surrounded by irrelevant visual stimuli. The difference between the two types of situation, which leads to interpretation of one context as distracting and another context as embedding, lies in the relationship of critical to irrelevant items. In the distraction situations, the figural properties of critical items remain intact. A triangle remains a triangle, despite the surroundings. In the embeddedness situation, the critical item or its parts are organized into new, competing gestalts which serve to break up the original critical figure. In effect, a distracting context may be thought of as obscuring a critical item without changing the nature of the item, whereas an embedding context serves to obscure a critical item because it changes the nature of the item.

### METHOD

Subjects were 150 undergraduate males attending Brooklyn College. All were paid volunteers. They ranged in age from 17 to 43 years (the second oldest was 27), in college class from lower freshmen to upper seniors, and included majors in virtually every subject offered at Brooklyn College.

The following tests were administered to each subject:

1. BAT. A measure of field dependence involving ability to adjust one's tilted body to the upright while surrounded by a tilted field. Score for each trial is degrees of deviation from the true vertical at which the subject reports himself as vertical. Test score is the sum of six trial scores.

2. RFT. In this test of field dependence the subject is seated in a completely darkened room, where all he can see is a luminous rod surrounded by a luminous square frame. At the beginning of each trial both rod and frame are presented in tilted positions, the subject's task being to adjust the rod to the vertical while the frame remains tilted. Score for each trial is degrees of deviation of the rod from the true vertical when the subject reports it to be upright. Test score is the sum of eight trial scores.

3. EFT. This test of field dependence consists of a series of pairs of figures, each pair consisting of a simple geometric figure and a complex colored figure within which the simple figure is embedded. The subject's task is to quickly locate the simple figure within the complex one, his score being

the time taken to do so. Test score is the sum of time scores for 12 trials.

4. Wechsler Adult Intelligence Scale (WAIS). Six subtests of the WAIS, Block Design, Object Assembly, Comprehension, Vocabulary, Arithmetic, and Digit Span, were administered and scored in the standard manner (Wechsler, 1958).

5. Match Problems. In this test of Adaptive Flexibility, the subject is shown a diagram of a lattice of squares which are formed by an arrangement of matchsticks. His task is to show how, by removal of a specified number of matchsticks, one can arrive at a given number of squares. Score is the number of correct solutions in 14 minutes.

6. Insight Problems. The test, another measure of Adaptive Flexibility, consists of 12 problems requiring the restructuring of the "given" in order to reach a solution. Score is the number of correct solutions in 22 minutes.

7. Digit Symbol. This modified version of the WAIS subtest contains approximately twice as many items as the standard version. Score is the time taken to complete the test.

8. Distracting Contexts Test I (DCT I). This test is similar to the EFT; however, the context in which the simple figure is hidden was designed to serve a distracting, rather than an embedding, function.<sup>3</sup>

9. DCT IIA. This test is scored for speed with which the subject locates each of a series of 13 simple geometric figures which are interspersed among a matrix of 52 such figures. Here, distraction is provided by the presence of irrelevant simple figures surrounding the critical ones.

10. DCT IIB. This test is similar to DCT IIA; however, as an additional source of distraction, a series of semitransparent colored designs are superimposed over many of the simple geometric figures to be located.

11. Arithmetic Operations Test. The subject's score is the time it takes to solve 24 simple arithmetic problems (e.g.,  $8 - 3 + 4$ ). The problems are spaced evenly on a single page; however, surrounding the problems are a series of totally irrelevant pictures, commands, sayings, and jokes. The subject is instructed to do the problems as quickly as he can, ignoring the irrelevant material.

In addition, two other tests (Form-Recognition and Cancellation), administered to the present subjects by another investigator as part of a different study of embeddedness and distraction (Loeff, 1961), were included in the present study without hypotheses for purposes of exploring their factorial content. The former test (Holtzman, 1955) involves speed of recognition of a series of impoverished pictures of animals, and the latter involves speed of crossing out the letters "a, t, and c" each time they appear on a page of randomly arranged letters.

The hypotheses of the present study include a general hypothesis, that tests involving overcoming

<sup>3</sup> Copies of the four distraction tests are available at no cost from Cognitive Tests, P. O. Box 4, Vanderveer Station, Brooklyn 10, New York.



TABLE 1  
PREDICTED FACTORS AND TEST LOADINGS

Factor	Name	Tests
I	Analytical Ability	BAT, RFT, EFT, Block Design, Object Assembly, Match Problems, Insight Problems
II	Verbal Comprehension	Comprehension, Vocabulary
III	Memory-Attention	Arithmetic, Digit Span, Arithmetic Operations
IV	Overcoming Distracting Contexts I	DCT I, DCT IIA, DCT IIB, EFT, Block Design, Object Assembly
V	Overcoming Distracting Contexts II	Arithmetic Operations, DCT I, DCT IIA, DCT IIB
VI	Overcoming Embedding Contexts versus Time	EFT, Block Design, Object Assembly, Match Problems, Insight Problems
VII	Performance Speed	Digit Symbol
VIII	DCT II Equipment	DCT IIA, DCT IIB

embeddedness (EFT, RFT, BAT, Block Design, Object Assembly, Match Problems, Insight Problems) will load a different factor (or factors) than tests involving overcoming distraction (DCT I, IIA, IIB, Arithmetic Operations); and a series of specific hypotheses as to the nature of factors expected to emerge and the tests expected to load such factors (summarized in Table 1).

Postulation of Factors I, II, and III was based upon the Goodenough and Karp study which obtained similar factors with children. Factor VII was previously identified by Cohen (1957) in studies of the WAIS. Factors IV, V, and VIII were expected as a result of the inclusion of the distraction tests, and Factor VI because of inclusion of the Adaptive Flexibility measures.

Pearson product-moment coefficients of correlation were obtained between each pair of the 18 measures. Principal components factors (Hotelling, 1933) were then extracted from the matrix of intercorrelations by computer.

The decision about number of factors to rotate was based upon two criteria, a rule-of-thumb estimate involving rotation of factors with latent roots of 1.00 or greater (Harman, 1960) and an estimate of Saunders' test (Cattell, 1952). The former criterion indicated that four, and the latter that eight, factors should be rotated. In view of this divergence, both the first four and the first eight principal components factors were separately rotated to oblique simple structure, following procedures outlined by Cattell (1952).<sup>4</sup> Rotation was carried out

<sup>4</sup> Five tables, giving zero-order intercorrelations, reliability estimates, unrotated factor loadings, and oblique transformation matrices have been deposited with the American Documentation Institute. Order Document No 7507 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in ad-

graphically by the present investigator while "blind" as to the identities of the test loadings being rotated. After rotation was completed the matrix of rotated factors was given to an independent judge (also "blind") to attempt further rotations that would improve simple structure.

For purposes of evaluation of hypotheses, loadings of  $\pm .25$  or greater were considered significant.

## RESULTS

Final rotated loadings of the 18 measures on four and eight factors are presented in Tables 2 and 3 and correlations among reference vectors in Table 4. In the course of carrying out rotations it quickly became apparent to the present investigator and to two independent judges that the four-factor solution underestimated the number of interpretable factors present. These factors were crowded and many test loadings tended to

TABLE 2  
FACTOR LOADINGS AFTER ROTATION OF FOUR FACTORS

Test	Factor			
	1	2	3	4
EFT	.82	-.06	-.17	.00
WAIS Arithmetic	.34	.04	.03	.57
WAIS Object Assembly	.59	.07	.09	-.18
Match Problems	.51	.10	.05	.10
DCT IIA	.14	.46	-.20	.05
Cancellation	-.10	.68	.10	.10
RFT	.78	-.10	-.33	.06
WAIS Vocabulary	.10	-.10	.64	.09
WAIS Block Design	.66	.03	.05	-.01
Arithmetic Operations	.00	.56	.01	.55
DCT I	.09	.56	.01	-.10
Form-Recognition	.44	.13	.09	-.32
BAT	.74	-.21	-.27	.07
WAIS Comprehension	-.07	.02	.76	-.01
WAIS Digit Span	.09	.10	-.09	.78
Digit Symbol	-.08	.75	-.01	.41
Insight Problems	.52	.00	.17	.05
DCT IIB	.05	.64	-.10	.08

Note.—Decimals are omitted; loadings over  $\pm .24$  are in boldface type; and tests are reflected where necessary so that positive scores represent greater field independence and intelligence, more rapid problem solving, letter cancellation, and form recognition, and less interference from distracting stimuli.

vance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Supplementary analytic rotations, to normal vari- max and bi-quartimin criteria, were also carried out. Results of these procedures, which are fairly consistent with those to be presented here, are reported elsewhere (Karp, 1962).



TABLE 3  
FACTOR LOADINGS AFTER ROTATION OF EIGHT FACTORS

Test	Factor							
	1	2	3	4	5	6	7	8
EFT	.73	-.09	-.04	.00	.01	.27	.09	.05
WAIS Arithmetic	.09	-.07	.09	.60	.17	.01	.09	-.01
WAIS Object Assembly	.49	-.05	.10	-.09	.10	.27	-.08	-.10
Match Problems	.59	.10	-.03	.01	-.23	.55	-.10	.10
DCT IIA	-.10	.08	-.03	.06	.62	-.09	-.10	.07
Cancellation	-.02	.56	.22	-.06	.09	.05	-.40	.04
RFT	.76	.05	-.05	.07	-.09	-.04	.00	-.03
WAIS Vocabulary	.08	.00	.76	-.07	-.10	.03	.04	-.05
WAIS Block Design	.56	-.09	-.01	.08	.04	.39	.07	-.07
Arithmetic Operations	.10	.61	.09	.49	-.10	-.03	-.47	-.09
DCT I	-.01	.48	.07	.02	.04	-.10	.01	-.24
Form-Recognition	.38	.07	.10	-.10	.08	.05	.05	-.30
BAT	.74	-.10	.01	.07	.03	-.05	-.08	.00
WAIS Comprehension	-.09	.02	.70	.08	.04	-.03	-.08	-.33
WAIS Digit Span	-.09	.08	-.01	.73	.05	-.09	-.02	.07
Digit Symbol	-.06	.61	-.07	.43	.06	.10	-.41	-.10
Insight Problems	.47	-.10	.02	.05	-.10	.56	.09	.03
DCT IIB	-.05	.52	.07	-.06	.07	.01	-.03	.09

Note.—See note to Table 2.

“hang” in a position between the hyperplanes and clusters of measures clearly loaded (e.g., 10% of the loadings fell between .10 and .25). Percentage of loadings within the  $\pm .10$  hyperplanes (57%) was relatively low for the kinds of ability variables included in the present study. In contrast, the eight-factor solution suggested considerably better resolution of factors, with 76% of loadings between  $\pm .10$  and only 2% between .10 and .25. Regarding the general hypothesis, that tests involving the overcoming of embedding

contexts load different factors than tests involving distracting contexts, it may be noted that there is a complete absence of overlap of significant loadings of tests involving different kinds of contexts on either four or eight factor solutions. None of the factors obtained load both kinds of tests. It may also be noted from Table 4 that Factors 1 and 6 (defined by embeddedness tests) are not highly related to any of the factors which load distraction tests (2, 4, 5, 7, 8), although some obliqueness is present (correlations between reference vectors ranging from .04 to  $-.27$ ). It seems reasonable to view these data as providing substantial support for the general hypothesis, with the reservation that factors defined by the embeddedness tests are not orthogonal to all factors defined by the distraction tests.

With regard to specific hypotheses about the kinds of factors expected and the tests predicted as loading such factors, a series of point-biserial correlations was obtained between each combination of one predicted and one obtained factor. For this purpose, predicted factors were treated as dichotomous variables (each test scored “1” if significant loading on that factor was predicted and “0” if not) and obtained factors as continu-

TABLE 4  
CORRELATIONS AMONG REFERENCE VECTORS AFTER  
OBLIQUE ROTATION OF EIGHT AND FOUR  
(IN ITALICS) FACTORS

	2	3	4	5	6	7	8
1	-.23 -.16	-.22 -.35	-.12 -.04	-.28	.35	.04	.06
2		.10 .00	.21 .33	-.12	-.16	-.43	-.08
3			-.18 -.18	.03	-.21	-.05	-.19
4				.11	-.16	-.18	-.09
5					-.27	.01	-.01
6						.05	.18
7							.02

ous variables (score for each test was its loading on a given factor). An obtained factor was matched with the predicted factor with which it was most highly correlated. The magnitude of the correlation suggests the degree to which the match is a good one.

Obtained Factor 1 appears highly related to Predicted Factor I (Analytic Ability) with  $r_{pb} = .90$  and  $.93$  ( $p < .001$ ) for four- and eight-factor solutions. Two tests that were not predicted as loading this factor, Arithmetic and Form-Recognition, significantly load it in the four-factor solution. Form-Recognition also loads the eight-factor solution.

Factor 2 is most highly related to Predicted Factor V (Overcoming Distracting Contexts II) with  $r_{pb} = .71$  and  $.64$  ( $p < .005$ ) for the two solutions. Three of the four distracting contexts tests (DCT I, IIB, Arithmetic Operations) load this factor in both analyses. The fourth distraction measure, DCT IIA, loads only the four-factor solution. In addition, Digit Symbol and Cancellation significantly load Factor 2 in both solutions, contrary to predictions.

Obtained Factor 3 relates most highly (.94 and .85;  $p < .001$ ) to Predicted Factor II (Verbal Comprehension), loading WAIS Vocabulary and Comprehension. No other tests were expected to load Factor II; however, low negative loadings on Factor 3 of RFT ( $-.33$ ) and BAT ( $-.27$ ) were obtained in the four-factor solution.

Obtained Factor 4 relates significantly to Predicted Factor III (Memory or Attention-Concentration), with  $r_{pb} = .85$  ( $p < .001$ ) for each solution. This factor loads Arithmetic, Digit Span, and Arithmetic Operations, as predicted. In addition, Digit Symbol loads Factor 4 in each solution.

Factors 5 through 8 were obtained only in the eight-factor solution. Factor 5 is most highly related to Predicted Factor VIII (DCT II Equipment), with  $r_{pb} = .63$  ( $p < .005$ ). Loading of DCT IIA and IIB on this factor was predicted; however, only the former occurs.

Factor 6 matches Predicted Factor VI (Overcoming Embeddedness versus Time), with  $r_{pb} = .89$  ( $p < .001$ ). This factor loads all embeddedness tests that are timed.

Factor 7, which loads Digit Symbol, Cancellation, and Arithmetic Operations, appears to correspond most closely to Predicted Factor VII (Performance Speed) which only postulated loading of Digit Symbol. For this comparison, a point-biserial coefficient of correlation is not appropriate as only one test was predicted as loading Factor VII. As a rough estimate of relationship between Predicted Factor VII and each of the obtained factors, loadings of Digit Symbol on the latter were converted to standard scores (based upon the distribution of loadings on each factor). The highest standard score for Digit Symbol occurred on Obtained Factor 7, which was on this basis matched with Predicted Factor VII.<sup>5</sup>

Factor 8 related significantly ( $r_{pb} = .48$ ,  $p < .05$ ) only to Predicted Factor II; however, the low level of this relationship and the nature of the three tests loading this factor (DCT I, Form-Recognition, and Comprehension) suggest that Factor 8 might best be viewed as a residual factor.

## DISCUSSION

The results thus provide support both for the general hypothesis and with regard to the character of four of the eight predicted factors (I, II, III, and V). Partial support may be found for two other predicted factors (VI and VII). Tests involving ability to overcome embeddedness clearly define Factor 1 and suggest a high degree of overlap between measures of field dependence, WAIS "closure" measures, and measures of Adaptive Flexibility based upon this common characteristic.

Measures reflecting ability to overcome the effects of distracting contexts contribute to the definition of Factor 2. In addition, loading of Digit Symbol and Cancellation is substantial. This might suggest either of two kinds of possibilities, that Digit Symbol and Cancellation involve distracting contexts or that this factor might tap other kinds of abilities than the overcoming of distraction. One might, on an ad hoc basis, interpret these

<sup>5</sup> If one were to assume a normal distribution of test loadings, loading of Digit Symbol of .41 or greater on Factor 7 could be expected to occur by chance alone in 1 out of 50 cases.



tests as involving routine clerical performance where items that must be found and manipulated are located within a matrix of generally similar irrelevant items. For Cancellation, the irrelevant (and thus, by implication, potentially distracting) background is provided by the randomly arranged matrix of letters among which the critical letters must be found. Digit Symbol involves, at least for the period of time until symbol-number pairs are learned, the successive matching of a series of test numbers with the same numbers in the legend. The legend provides an orderly matrix of similar items (number-symbol pairs), the correct one of which must be selected for each test item. Thus, one might consider Factor 2 to represent ability to rapidly select critical items from a matrix of similar irrelevant items. In this view, the degree to which one is distracted by the irrelevant items might affect his speed of doing the task.

Factors 3 and 4 closely match the Verbal Comprehension and Memory factors obtained by many investigators using one or another of the Wechsler scales. Loading of Digit Symbol on Factor 4 was not predicted, on the basis of results by Cohen (1957, 1959) with the WISC and WAIS; however, such loading agrees with the results of Goodenough and Karp (1961), using the WISC with children, where Coding (the WISC version of Digit Symbol) loads their Attention-Concentration factor.

The choice of a name for Factor 4 has been a subject of some discussion in the literature, the most frequently used choices involving either short-range memory (as memory for digits over an interval of a few seconds) or some combination of attention-distraction-concentration. Yet, the alternative interpretations may not be very far apart. Undoubtedly, tests such as WAIS Arithmetic and Digit Span involve memory. However, it also seems reasonable to consider inability to attend as one of the most likely reasons for failure of short-range memory. The present investigator would favor representation of both interpretations in the name (e.g., Memory-Attention) until such time as further evidence suggests a review of this issue.

Factor 5 loads only DCT IIA, an outcome neither predicted in advance nor clear at this time.

Factor 6 (Overcoming Embeddedness Against Time) loads all timed embeddedness tests as predicted. However, those tests which provide the least speed pressure (Match and Insight Problems) are most highly loaded on this factor, whereas the remaining embeddedness tests (EFT, Object Assembly, Block Design) have relatively small loadings. On this basis one might consider an alternative interpretation of this factor.

Both Match and Insight Problems define the Guilford et al. (1957) Adaptive Flexibility factor which, as has been suggested above, may be highly similar to Analytic Ability. In fact, the presence of these tests on Factor 1 would support this view. However, it is also possible that these tests and, by implication, the Adaptive Flexibility factor involve other kinds of abilities than overcoming embeddedness. One could, for example, postulate a kind of persistence-at-a-difficult-task ability or a speed-at-solving-complex-problems ability to account for Factor 6 variance. Such abilities might be required to a greater degree by Match and Insight Problems than by the other tests which load Factor 6. To evaluate such hypotheses it would be necessary to include, in a single study, measures of persistence or problem solving ability (or any other postulated ability) which do not also involve overcoming embeddedness.

Factor 7 (Performance Speed) loads Cancellation, Arithmetic Operations, and Digit Symbol. The presence of Cancellation on this factor would tend to emphasize the routine clerical performance aspect of tests loading this factor as opposed, for example, to the distraction and memory aspects of Arithmetic Operations and Digit Symbol (which have been accounted for in interpretation of other factors which load these tests). The presence on this factor of two tests that are not subtests of the WAIS suggests that this may be the Digit Symbol specific factor obtained by Cohen (1957, 1959) in his analyses of the WAIS and WISC.



There would also seem to be overlap between Factor 7 and a Perceptual Speed factor identified in several studies. Davis (1956) and Pemberton (1952) describe this factor as involving speed of identification of a given stimulus within a mass of distractors, where the task is a simple one and where the distractors do not serve to distort or conceal the stimulus, as through embeddedness. This description would seem appropriate for two of the three tests which define Factor 7, Digit Symbol and Cancellation. It is not as relevant for Arithmetic Operations, where rapid location of the critical items (arithmetic problems) would seem to represent a relatively minor aspect of the task. Success at Arithmetic Operations would appear to involve ability to attend to the problems to be solved in the face of distraction, rather than ability to locate the problems among a matrix of irrelevant stimuli. Here, the irrelevant materials might be seen as "competing" with the critical problems for the attention of the subject. It is possible to advance a similar kind of interpretation for the presence of Digit Symbol and Cancellation on this factor, if we view the tasks involved as boring or tiresome because of their repetitiousness. Here, other kinds of stimuli which have not been intentionally introduced into the test situation (e.g., outside sounds, ideas which occur to the subject) may compete with the task for the attention of the subject. He might be distracted not because the distractors are particularly enticing, but because the critical task is so uninteresting.

Factor 8 corresponds to no predicted factor, and would appear to be a residual factor. Similarly, Predicted Factor IV does not relate highly to any obtained factor.

Considering all seven interpretable factors obtained in the present study, it may be noted that four of these factors have been viewed as involving, in some way, attention, concentration, or distraction. Factor 2 is seen as reflecting ability to overcome the effects of distracting (perceptual) contexts, Factor 4 as involving attention as an aid to short-range memory, Factor 5 as reflecting specific variance of Distracting Contexts Test IIA, and Factor 7 as loading measures of Per-

formance Speed, which would appear to require some freedom from distraction. One might, in addition, consider Factor 6, involving all timed embeddedness tests, as potentially influenced by distraction. Yet, each of these factors appears quite distinct and is not highly correlated with any other factor. If nothing else, the presence of so many factors which directly or indirectly may be affected by some form of distraction would suggest that there may be no consistent kind of ability to attend or resist distraction which cuts across a great variety of situations and sources of distraction. This possibility is even more compelling when one considers that the college students of the present sample are, for the most part, above average in general intelligence and have demonstrated at least moderate scholastic success, which would seem to entail some kind of ability to attend and concentrate.

In general, the results would favor the position of Witkin and his colleagues, that measures of field dependence involve ability to overcome the effects of embedding contexts and that other kinds of cognitive tasks involving this ability load the same factor as the measures of field dependence. The results further suggest that ability to overcome embeddedness is factorially different from ability to resist distraction, although some correlation occurs between factors representing these abilities.

#### REFERENCES

- BOTZUM, W. A. A factorial study of the reasoning and closure factors. *Psychometrika*, 1951, 16, 361-386.
- CATTELL, R. B. *Factor analysis: An introduction and manual for the psychologist and social scientist*. New York: Harper, 1952.
- COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, 21, 283-290.
- COHEN, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *J. consult. Psychol.*, 1959, 23, 285-299.
- DAVIS, P. C. A factor analysis of the Wechsler-Bellevue Scale. *Educ. psychol. Measmt.*, 1956, 16, 127-146.
- GARDNER, R. W., JACKSON, D. N., & MESSICK, S. J. Personality organization in cognitive controls and intellectual abilities. *Psychol. Issues*, 1960, 2 (Whole No. 8).

- GOODENOUGH, D. R., & KARP, S. A. Field dependence and intellectual functioning. *J. abnorm. soc. Psychol.*, 1961, 63, 241-246.
- GUILFORD, J. P., FRICK, J. W., CHRISTENSEN, P. R., & MERRIFIELD, P. R. A factor-analytical study of flexibility in thinking. Laboratory Report No. 18, 1957, University of Southern California, Department of Psychology.
- HARMAN, H. H. *Modern factor analysis*. Chicago: Univer. Chicago Press, 1960.
- HOLTZMAN, W. H. Tentative manual: The Holtzman Form-Recognition Test. Austin: University of Texas, 1955. (Mimeo)
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, 24, 417-441; 498-520.
- JACKSON, D. N. Stability in resistance to field forces. Unpublished doctoral dissertation, Purdue University, 1955.
- KARP, S. A. A factorial study of overcoming embeddedness in perceptual and intellectual functioning. Unpublished doctoral dissertation, New York University, 1962.
- LOEFF, R. G. Embedding and distracting field contexts as related to the field-dependence dimension. Unpublished master's thesis, Brooklyn College, 1961.
- PEMBERTON, CAROL L. The closure factors related to other cognitive processes. *Psychometrika*, 1952, 17, 267-288.
- PODELL, J. E. Personality and stimulus factors in adult cognition: A developmental analysis of de-contextualization. Unpublished doctoral dissertation, Clark University, 1957.
- PODELL, J. E., & PHILLIPS, L. A developmental analysis of cognition as observed in dimensions of Rorschach and objective test performance. *J. Pers.*, 1959, 27, 439-463.
- THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. Chicago Press, 1944.
- THURSTONE, L. L. Mechanical aptitude: III. Analysis of group tests. Report No. 55, 1949, University of Chicago, Psychometric Laboratory.
- WECHSLER, D. *Measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins, 1958.
- WITKIN, H. A., DYK, RUTH B., FATERSON, HANNA F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received October 1, 1962)

## COMPARATIVE EFFICIENCY OF SOME TESTS OF CEREBRAL DAMAGE<sup>1</sup>

MAURICE KORMAN AND STANLEY BLUMBERG

*University of Texas Southwestern Medical School*

This study: (a) compared 40 Ss with cerebral damage with 40 matched nondamaged Ss on 4 commonly used psychometric instruments, as well as W-B Vocabulary and MMPI L Scale; and (b) investigated the relationship between these tests and 4 dimensions (laterality, severity, progressiveness, and diffuseness). Using artificial base rates, the rank order for "accuracy" was: Memory for Designs Test (MFD), Spiral Aftereffect Test (SAET), Trail Making Test (TMT), Bender Gestalt, Vocabulary, L. Combining the 3 "best" tests increased accuracy slightly. Some variables were significantly associated with laterality (TMT, A minus B; MFD); severity (SAET); progressiveness (L; TMT, A minus B; age); diffuseness (age).

Problems connected with the assessment of cerebral damage (CD) continue to be among the more vexing that the clinician faces in his diagnostic activities. Although the literature is extensive, few studies, as Jenkins and Lykken (1957), and more recently Reitan (1962), have remarked, deal with more than one test at a time. This prevents comparisons of the relative discriminating power of various CD tests in a situation where other parameters, including base rates, are held constant. Furthermore, the independent variable of most studies has too often been merely "brain damage" with no attempt made to investigate relationships between various dimensions of cerebral damage and psychometric instruments. It can be argued that loss of significant variance can sometimes occur through the use of such overly global constructs (Korman, 1960).

The present study was undertaken as a pilot attempt to explore simultaneously the usefulness of certain commonly used CD tests in intergroup (CD versus non-CD) as well as intragroup (all CD) criterion situations. It must, however, be pointed out that such a design *cannot*, because of its artificiality,

yield results that have immediate applicability.

We were also interested in testing the suggestion of Garrett, Price, and Deabler (1957) that hit rate might be increased by the *joint* use of the more valid tests. We chose the tests in question because of their relatively widespread use among working clinicians. In some instances, furthermore, hypotheses concerning relationships between such tests and various CD dimensions have been formulated. This is particularly true of the Trail Making Test (Reitan & Tarshes, 1959), the Spiral Aftereffect Test (Sindberg, 1961) and the Memory-for-Designs Test (Graham & Kendall, 1960).

Finally, the intercorrelations among these tests were computed. Blau and Schaffer's (1960) data suggest a moderate relationship among somewhat similar instruments. In addition the relationship of such variables as age, education, and vocabulary to CD measures is of particular methodological importance.

### METHOD

#### *Subjects*

The CD group was composed of 52 patients from the in- and outpatient services of one county general hospital and two Veterans Administration hospitals. The only criteria for selection were that the patient have definite cerebral damage and that he be able to take the battery of tests. For each patient, a rating sheet patterned after one developed by Reitan (1959) was completed by the staff or resident neurologist assigned to the case. This rating sheet provided information as to diagnosis, type

<sup>1</sup> We wish to acknowledge the cooperation of the Veterans Administration Hospitals at Dallas and McKinney, Texas, the unfailing helpfulness of Sven Gustaf Eliasson, MD, and the assistance of Robert Via and Betty Jarrett in administering tests.

Preparation of this paper was facilitated by a General Research Support Grant (1 GS 129) from the National Institutes of Health.



of lesion, degree of severity, whether focal and/or diffuse, localization, degree of progressiveness, and disease process. In addition, the number and kind of diagnostic techniques used (e.g., EEG, skull X ray, pneumoencephalogram, etc.) and the degree of confidence in the diagnosis were rated by the neurologist. The median number of diagnostic tools used with each patient was four. Degree of confidence in diagnosis was "good" or "excellent" in 88% of the cases rated. The group was composed of 40 male and 12 female subjects, whose diagnoses fell into the following major categories: cerebral vascular disease, 16; tumor, 9; trauma, 7; epilepsy, 7; degenerative or demyelinating disease, 6; cerebral vascular disease plus trauma, 2; inflammatory or infectious disease, 1; frontal lobotomy, 1; multiple cortical lesions, 1; trauma plus tumor, 1; and narcolepsy, 1.

The distribution of patients according to the rated dimensions were as follows: cerebral laterality—18 left side, 16 right side, and 18 bilaterally damaged; severity—6 mild, 30 moderate, and 16 severe; progressiveness—33 static and 16 progressive; diffuseness—12 focal, 23 diffuse, and 17 focal and diffuse. It should be noted that because of the neurologists' uncertainty, 3 subjects were not rated on the static versus progressive dimension.

Of the 52 CD patients, 40 were matched in pairs for age, education, and sex with a control group of non-brain-damaged patients from the in- and outpatient psychiatric and medical services of the same hospitals from which the CD patients were selected. These 40 CD subjects had a mean age of 45.37 with an *SD* of 15.80. Their mean years of education was 9.80, *SD* 3.20. Their distribution by diagnosis was 12 cerebral vascular disease, 8 tumor, 6 epilepsy, 4 trauma, 4 degenerative or demyelinating disease, 2 CVD and trauma, 1 tumor and trauma, 1 inflammatory disease, 1 lobotomy, and 1 multiple cortical lesions. The 40 control subjects had a mean age of 45.10, *SD* 15.45, and

mean education of 9.83, *SD* 2.53. Having been selected from the same hospitals and clinics, as well as matched on age and education, the two groups were assumed to be of comparable pre-morbid intelligence. Twenty-four of the control subjects had psychiatric diagnoses as follows: schizophrenia, 8; psychoneurosis, 6; alcoholism, 5; and personality disorder, 5. The remaining 16 subjects had a variety of medical diagnoses such as duodenal ulcer, pancreatitis, etc. There was no evidence of CD symptoms or diagnosis in the medical charts of the 40 control subjects.

## Tests

Every subject was individually administered each test in the battery. The order of presentation was staggered. Each test was administered and scored in accordance with the standard instructions in the test manual, except as noted below. The tests were the Visual Motor Gestalt Test (Bender, 1946), scored according to the system proposed by Pascal and Suttell (1950); the Trail Making Test (Reitan, 1956), from which four separate raw time scores were analyzed—Part A, Part B, Total, and A minus B; the Graham-Kendall Memory-for-Designs Test (MFD) (Graham & Kendall, 1960), both uncorrected and corrected scores; the Spiral Aftereffect Test (SAET) (Psychological Research and Development Corporation, 1958), with 8 trials presented in ABBABBA order using a 7-inch disk rotated at a speed of 80 rpm; the *L* scale of the MMPI; and the Wechsler-Bellevue I Vocabulary test.

## RESULTS AND DISCUSSION

### Cerebral Damaged versus Control Subjects

Table 1 summarizes the performance of the CD and control groups. Although there are marked differences in the tests' effectiveness as reflected in their overall hit rate, the CD

TABLE 1  
SUMMARY DATA COMPARING 40 CEREBRAL DAMAGED AND 40 NONDAMAGED SUBJECTS

	Non-CD		CD		Cutting score	% CP <sup>a</sup> total	% CP <sup>a</sup> CD	% CP <sup>a</sup> non-CD
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Age	45.10	15.45	45.37	15.80				
Education	9.83	2.53	9.80	3.20				
MFD, uncorrected	3.15	3.70	11.00	6.53	5-6	90.00	90.00	90.00
MFD, corrected	.93	3.75	7.95	6.63	2-3	82.50	77.50	87.50
SAET	7.50	1.05	5.15	2.23	7-8	82.50	85.00	80.00
Trails B	87.30	52.95	179.95	102.60	107"-108"	82.50	87.50	77.50
Trails, Total	139.00	70.28	287.55	172.85	189"-190"	80.00	72.50	87.50
Trails A	51.70	26.85	107.60	79.05	71"-72"	76.25	63.50	90.00
Bender	37.78	22.55	82.05	49.00	44-45	73.75	77.50	70.00
Trails A-B	35.60	45.95	72.35	64.05	63"-64"	70.00	47.50	92.50
Vocabulary	20.93	7.80	15.90	7.36	14-15	67.50	57.50	77.50
<i>L</i>	5.05	2.25	6.15	2.60	6-7	63.75	47.50	80.00

<sup>a</sup> Correct predictions.

mean scores on all variables except age and education (on which the groups were matched) were significantly poorer ( $p < .05$  or better) than the control group's means. It cannot be emphasized too strongly that the cutting scores reported in Table 1 *must not be generalized to other populations and other problems*. For the purpose of comparing the efficiency of these tests to one another, highly artificial base rates of .50 CD and .50 non-CD were used. In clinical situations, where base rates are typically asymmetrical, it is obvious that different cutting scores would be optimal, and that these would shift depending on the acceptable proportions of false negatives and false positives (see Meehl & Rosen, 1955).

The most effective single test in correctly predicting the presence or absence of CD was the MFD, uncorrected for age and vocabulary. Its efficiency was equally great in predicting the absence as well as the presence of CD, the hit rate being 90% in both instances, using a cutting score between 5 and 6. These data compare fairly closely with the information given in the newest manual for the MFD. Recomputing the validation and cross-validation data of their Table 11 (Graham & Kendall, 1960, p. 162) we obtained a total hit rate of 84% using our cutting score. However, if we use Graham and Kendall's system of breaking down MFD scores into normal, borderline, and critical areas, only 32.5% of our CD group would be correctly identified, while 2.5% of that group would be misclassified. Although we are in general agreement with Graham and Kendall's conservative approach, which emphasizes interpretation of deviant scores only, we wonder whether cutting scores somewhere between 5 and 11 would not have the asset of lowering the false negative rate without undue violence to base rate problems. Our experience with this test suggests that the lenient scoring system likewise contributes to this difficulty.

The corrected MFD scores (i.e., the difference between subjects' actual scores and those expected on the basis of age and education) did not do quite so well as the uncorrected scores, the overall hit rate being 82.5%. It seemed at first that this might partly be due to the fact that our two groups were already carefully matched for age and

education. Table 4, however, suggests an additional answer. In contrast to Graham and Kendall, we find a negligible correlation ( $r = -.09$ ), between age and uncorrected MFD scores. Under these conditions correcting for age increases error variance. On the other hand, we find an appreciable correlation between MFD uncorrected scores and vocabulary ( $r = -.48$ ), which drops only slightly upon correction ( $r = -.34$ ). Aside from this negligible difference, perhaps MFD scores should not be corrected for vocabulary at all, inasmuch as vocabulary level may be less of an irrelevant variable than had been thought. These data seem to bolster the questions which Hunt (1952) has raised concerning the need for correcting MFD scores.

The SAET also yielded an overall hit rate of 82.5%. For our sample it seemed to do slightly better with true positives than with true negatives. Because investigators have dealt with the SAET more as a technique than as a test, thus bypassing the problems of standardized administration and scoring, comparisons are difficult to make. The hit rates of six studies reported in recent years range from 67.5% to 96.4%, with a median of 87.2%. In the present study, we utilized the preliminary manual provided by the Psychological Research and Development Corporation, Tampa, Florida, the makers of the apparatus. Now that methodological studies such as that of Sindberg (1961) are being published, it is to be hoped that this technique will take one more of the characteristics of a methodologically sound testing procedure. The distribution of scores in our samples tends to corroborate Sindberg's conclusion that a longer series of trials increases between-group discrimination. The majority of studies (see Blau & Schaffer, 1960), it should be recalled, have used four or fewer trials. In contrast to the significant correlation between SAET and age of  $-.39$  reported by Goldberg and Smith (1958), we find a correlation of only .08.

Analysis of the four different scores from the Trail Making Test indicates that Part B and Total Score differentiate the CD group from the controls better than do Part A or A minus B. As shown in Table 1, using optimum cutting scores, the total hit rate for Part B is 82.5%, for Total Score, 80%,



for Part A, 76.3%, and for A minus B, 70%. Our results are in general agreement with previous studies indicating that persons with brain damage tend to perform significantly more poorly on the Trail Making Test than do control subjects without brain damage (Armitage, 1946; Davids, Goldenberg, & Laufer, 1957; Reitan, 1955, 1958). An exception to this generalization is the study of Brown, Casey, Fisch, and Neuringer (1958) which found CD and psychotic groups significantly different from normals, neurotics, and character disorders, but not significantly different from each other. In our study, 8 control subjects had diagnoses of schizophrenia; the total control group was significantly superior to the CD group. In addition, separate analysis of the scores of the schizophrenic subjects revealed only two misclassifications (false positives) on Part A, Part B, and Total Score of the Trail Making Test.

A cutting score between 44 and 45 yielded a total hit rate of 73.8% on the Bender, with 77.5% of the CD group scoring 45 or above, and 70% of the controls scoring 44 or below. The group mean scores of 82 and 37.8 are comparable to those reported by Bowland and Deabler (1956). They obtained a mean of 80 from the CD subjects, 42.8 from the schizophrenics, 28.1 from neurotics, and 16.4 from normals. Our total hit rate of 73.8% is somewhat less than the 80% reported by Goldberg (1959).

Vocabulary tests have not been typically evaluated as tests of cerebral damage. On the contrary, instruments built around the differential score technique have implied a fundamental constancy of vocabulary performance regardless of brain changes. Such recent data as that of Williams, Lubin, and Gieseeking (1959) clearly suggest that this is not the case. Likewise, our CD group, which had been carefully matched for educational level, as well as age, with the non-CD group, nonetheless had significantly lower vocabulary scores. Whereas the control subjects had a mean vocabulary of 20.93, the CD subjects had a mean vocabulary of 15.90. A cutting score between 14 and 15 yielded correct predictions in 67.50% of all cases. However, the test's capacity for correctly identifying nondamaged patients is consider-

TABLE 2  
PREDICTIONS OF CEREBRAL DAMAGE BY VARIOUS  
COMBINATIONS OF THE THREE "BEST" TESTS  
(MFD, TRAILS B, SAET)

Number of tests with "positive" findings	CD subjects ( <i>N</i> = 40)	Non-CD subjects ( <i>N</i> = 40)
3	25	2
2	12	2
1	2	10
0	1	26

ably higher than its efficiency for detecting CD. It would seem that tests which utilize vocabulary scores as correction factors may unwittingly be blunting the sharpness of their discriminatory power. It is interesting to note in Table 4 that vocabulary relates significantly to all the measures of CD of this study, the median *r* being .33.

The MMPI *L* scale has the lowest discriminating power of any of our variables. Its total correct hits are only 63.75%, and this at the price of over 50% false negatives. Its low discriminating power is not surprising, although our results do lend some support to Doehring and Reitan's (1960) notions about the CD patient's need to maintain as rigidly acceptable a self-concept as he can.

Table 2 reports our attempt to test the efficacy of the three most discriminating tests that were also independent when used jointly. The three measures are MFD, Trails B, and SAET. All of our subjects' records were examined for the number of tests indicating cerebral damage using the cutting scores of Table 1. Table 2 indicates that, using a cutting score of two or more tests with "positive" results, 37 or 92.5% of the CD subjects are correctly classified, while only four or 10% of the control subjects are inappropriately labeled damaged. The total hit rate of 91.75% represents a slight increment over our most successful test, the MFD, which by itself correctly classified 90% of our subjects. Because of the larger and more diverse sample of behavior upon which this index rests, however, it would probably show less shrinkage upon cross-validation than selection ratios based on single tests.



*Dimensions of Cerebral Damage*

One-way analyses of variance, or *t* tests when appropriate, were run on all the variables of this study including age and education with the CD subjects categorized with relevance to cerebral laterality, severity, progressiveness, and diffuseness. In cases where the results exceeded the .05 level of significance, degree of overlap and optimum cutting scores were determined. The interrelationships of these four criterion dimensions were investigated by means of chi square. In only one instance did the results approach significance. Ratings of diffuseness were related to bilaterality ( $p < .10$ ).

Three measures significantly differentiated the CD subjects according to laterality of damage. As indicated in Table 3, a cutting score between 59 and 60 seconds on the Trail Making A minus B score yielded a total correct classification of 74% of the 52 brain damaged patients. Those patients with bilateral damage performed similarly to those with left-sided damage on the Trail Making Test. Seventy-two percent of the bilaterally damaged patients and 72% of the left-side damaged patients had A minus B scores of 60 seconds or more, whereas 79% of the right-side damaged patients had scores of 59 seconds or less.

There are both similarities and differences between our results and those obtained by

Reitan and Tarshes (1959) in the only other published study relating Trail Making Test scores to laterality of brain lesions. Using converted *T* scores instead of raw time scores, Reitan and Tarshes obtained a 75% hit rate for left-side patients (compared to our 72%) and 72% hits for right-side subjects (compared to our 79%). In both studies the mean difference scores between the left- and right-side damaged patients were significantly different, with the left-side damaged patients performing worse on Part B in relation to Part A. In contrast to the similarity between the bilaterally damaged group and the left-side group in our study, Reitan and Tarshes (1959) found a significant difference between subjects with bilateral and those with left-side damage, but no significant difference between bilaterally and right-side damaged patients.

The MFD scores significantly differentiated laterality among the brain-damaged patients with higher hit rates for the bilaterally damaged patients than for those with either left- or right-side damage. As shown in Table 3, the optimum cutting scores result in 74% and 70% total correct classification for MFD uncorrected and corrected scores, respectively. Graham and Kendall (1960) hypothesized that the MFD test would be especially difficult for patients with lesions of the right hemisphere, but found that the test did not

TABLE 3  
SIGNIFICANT RELATIONSHIPS BETWEEN VARIABLES AND DIMENSIONS OF CEREBRAL DAMAGE

Dimensions and tests	Cutting score	% Correct predictions
<b>Laterality</b>		
1. Trails A-B	59"-60"	Total: 74% (Lt 72%; Rt 79%; Bil 72%)
2. MFD, uncorrected	9-10	Total: 74% (Lt 67%; Rt 64%; Bil 89%)
3. MFD, corrected	5-6	Total: 70% (Lt 61%; Rt 64%; Bil 83%)
<b>Severity</b>		
1. SAET	5-6	Total: 67% (Severe, 69%; Moderate, 67%)
<b>Static-Progressive</b>		
1. L	3-4	Total: 74% (Static, 91%; Progressive, 38%)
2. Trails A-B	55"-56"	Total: 67% (Static, 70%; Progressive, 63%)
3. Age	46-47	Total: 59% (Static, 64%; Progressive, 50%)
<b>Diffuseness</b>		
1. Age	52-53	Total: 73% (Focal, 83%; Diffuse, 70%; Focal and Diffuse, 70%)

\* Mild cases were excluded because of small *N* (*N* = 6).

differentiate right- from left-side damaged patients. Our bilaterally damaged patients and those with lesions on the left side performed worse (had higher scores) than did those with right-sided damage. However, our hit rates in identifying left- and right-sided lesions are not much above chance (in the low to middle 60s), while the rate for correctly identifying bilaterally damaged patients is 89% and 83% for uncorrected and corrected scores respectively. The differences between the three CD groups classified as to laterality cannot be attributed to age or education.

The dimension of severity, rated in terms of extent of disability, was discriminated only by the SAET. Because only 6 mild cases were available, they were excluded from the analysis. The 16 severe cases tended to report seeing the aftereffect less frequently than did the 30 moderate cases. With a cutting score between 5 and 6, correct predictions could be made with nearly the same frequency for both of these groups, yielding an overall hit rate of 67%. Although the separation is only moderately accurate, these data corroborate the report by Sindberg suggesting that while location of damage is unimportant with reference to SAET performance, "extensive or severe damage to any part of the cortex markedly reduces the probability of SAET report [Sindberg, 1961, p. 136]."

The neurologist's rating of the progres-

siveness of the disease process was dichotomized into static versus progressive. The 33 patients with static lesions tended to get higher scores on the *L* scale and on Trails A minus B, and to be older. As Table 3 indicates, the *L* scale and Trails A minus B identified correctly 74% and 67% of the total sample, respectively, while age identified only 59% of the subjects correctly. This dimension of progressiveness is confounded with disease process. For example, of 15 patients with cerebral vascular disease, 14 were rated as static; all 9 tumor cases were rated progressive; and all of the trauma cases were rated static. The 6 cases of degenerative or demyelinating disease were evenly divided between static and progressive.

Only age distinguished focal from diffuse lesions with the diffuse group showing a tendency to be older than either of the other two groups. A cutting score between 52 and 53 correctly identified 73% of the total CD group. All of the subjects with degenerative or demyelinating disease and most of those with cerebral vascular disease were rated as diffuse. The relationship of age to diffuseness apparently reflects the propensity of older people to suffer from disorders of a degenerative or arteriosclerotic type which tend to lack a clear focus.

Table 4 shows the magnitude of the correlations among the variables of this study based on an *N* of 80 (40 CD and 40 non-damaged subjects).

TABLE 4  
INTERCORRELATION MATRIX

	SAET	Trails A	Trails B	Trails, Total	Trails A-B	Bender	MFD un- cor- rected	MFD cor- rected	Vocab- ulary	Educa- tion	Age
<i>L</i>	-.15	.10	.10	.11	.06	.05	.38	.06	-.23	-.01	.18
SAET		.36	.37	.40	.20	.35	.40	.39	.32	-.02	.08
Trails A			.78	.92	.13	.05	.37	.30	-.30	-.33	-.01
Trails B				.96	.72	.33	.43	.39	-.38	-.25	-.12
Trails, Total					.50	.38	.43	.37	-.42	-.30	-.08
Trails A-B						.05	.27	.28	-.28	-.03	-.14
Bender							.60	.57	-.30	-.10	-.09
MFD, uncorrected								.96	-.48	-.30	-.09
MFD, corrected									-.34	-.21	-.31
Vocabulary										.49	.11
Education											-.07

Note.—.22 significant at .05 level, .28 significant at .01 level.



Of the 45 correlations among the 10 test scores analyzed (including Vocabulary and L), 32 are significant at the .01 level. The three most promising independent measures (MFD, uncorrected, SAET, and Trails B) have a mean intercorrelation of .40. Although these variables seem to share to some extent in a common underlying factor it is clear, from the fact that only some 16% of the variance is accounted for, that they tap relatively discrete abilities, even though they are nearly equally effective in discriminating the CD from the non-CD subjects.

The mean correlation between vocabulary and the other nine measures is .34. This association seems to reflect the fact, discussed above, that vocabulary apparently decreases to some extent with cerebral damage. By contrast, education is less closely related to the measures of this study (mean  $r = .17$ ), as is age (mean  $r = .12$ ). The latter, in particular, is a surprising finding; it reoccurs quite consistently, however, as witnessed by the fact that only one out of 9 correlations is significant at the .05 level.

Although the dimensions of brain damage investigated here are undoubtedly too gross (e.g., all patients with lesions in the left hemisphere regardless of the lobe involved probably cannot be expected to perform similarly), a beginning was made in relating some commonly used tests of cerebral damage to dimensions somewhat less global and heterogeneous than brain damage.

#### REFERENCES

- ARMITAGE, S. G. An analysis of certain psychological tests used for the evaluation of brain injury. *Psychol. Monogr.*, 1946, 60(1, Whole No. 277).
- BENDER, L. *Instructions for the use of Visual Motor Gestalt Test*. New York: American Orthopsychiatric Association, 1946.
- BLAU, T. H., & SCHAEFFER, R. E. The Spiral Aftereffect Test as a predictor of normal and abnormal EEG records in children. *J. consult. Psychol.*, 1960, 24, 35-42.
- BOWLAND, J. A., & DEABLER, H. L. A Bender-Gestalt diagnostic validity study. *J. clin. Psychol.*, 1956, 12, 82-84.
- BROWN, E. C., CASEY, A., FISCH, R. I., & NEURINGER, C. Trail Making Test as a screening device for the detection of brain damage. *J. consult. Psychol.*, 1958, 22, 469-474.
- DAVIDS, A., GOLDENBERG, L., & LAUFER, M. W. The relation of the Archimedes Spiral Aftereffect and the Trail Making Test to brain damage in children. *J. consult. Psychol.*, 1957, 21, 429-433.
- DOEHRING, D. G., & REITAN, R. M. MMPI performance of aphasic and nonaphasic brain-damaged patients. *J. clin. Psychol.*, 1960, 16, 307-309.
- GARRETT, E. S., PRICE, A. C., & DEABLER, H. L. Diagnostic testing for cortical brain impairment. *Arch. neurol. Psychiat.*, 1957, 77, 223-225.
- GOLDBERG, L. R. The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *J. consult. Psychol.*, 1959, 23, 25-33.
- GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, 22, 153-157.
- GRAHAM, F. K., & KENDALL, B. S. *Memory-for Designs Test: Revised general manual. Percept. mot. Skills*, 1960, 11, 147-188. (Monogr. Suppl. No. 2-VII)
- HUNT, H. F. Testing for psychological deficit. In D. Brower & L. E. Abt (Eds.), *Progress in clinical psychology*. Vol. 1. New York: Grune & Stratton, 1952.
- JENKINS, J. J., & LYKKEEN, D. T. Individual differences. *Annu. Rev. Psychol.*, 1957, 8, 79-112.
- KORMAN, M. Some problems in the psychological diagnosis of brain damage: An overview. *Texas Rep. bio. Med.*, 1960, 18, 214-221.
- MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, 52, 194-216.
- PASCAL, G. R., & SUTTELL, B. J. *The Bender-Gestalt Test: Its quantification and validity for adults; with scoring manual*. New York: Grune & Stratton, 1950.
- PSYCHOLOGICAL RESEARCH AND DEVELOPMENT CORPORATION. *Spiral aftereffect manual*. Tampa, Fla.: Author, 1958.
- REITAN, R. M. The relation of the Trail Making Test to organic brain damage. *J. consult. Psychol.*, 1955, 19, 393-394.
- REITAN, R. M. *Trail Making Test: Manual for administration, scoring, and interpretation*. Indianapolis, Ind.: 1956. (Mimeo)
- REITAN, R. M. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept. mot. Skills*, 1958, 8, 271-276.
- REITAN, R. M. The effects of brain lesions on adaptive abilities in human beings. Indianapolis, Ind.: 1959. (Mimeo)
- REITAN, R. M. Psychological deficit. *Annu. Rev. Psychol.*, 1962, 13, 415-444.
- REITAN, R. M., & TARSHES, E. L. Differential effects of lateralized brain lesions on the Trail Making Test. *J. nerv. ment. Dis.*, 1959, 129, 257-262.
- SINDBERG, R. M. Some effects of stimulus variation on spiral aftereffect in organic and non-organic subjects. *J. consult. Psychol.*, 1961, 25, 129-136.
- WILLIAMS, H. L., LUBIN, A., & GIESEKING, C. F. Direct measurements of cognitive deficit in brain injured patients. *J. consult. Psychol.*, 1959, 23, 300-305.

(Received October 10, 1962)



## PATIENTS AND THERAPISTS ASSESS THE SAME PSYCHOTHERAPY<sup>1</sup>

HERMAN FEIFEL AND JANET EELLS

*Veterans Administration Outpatient Clinic, Los Angeles*

This study analyzed the perceptions of both patients and their psychotherapists at the close of psychotherapy as to changes taking place and ideas about what was helpful and not helpful. The participants consisted of 63 outpatients and 28 psychotherapists. Data were secured by administering an open-end questionnaire. Major findings and conclusions were: (a) therapists stressed changes in symptomatic relief and improvement in social relationships, whereas patients focused on self-understanding and self-confidence; (b) patients underlined the opportunity to talk over problems and the "human" characteristics of the psychotherapist as helpful, and therapists highlighted therapeutic technique and support to the patient as most beneficial; (c) expectancy and conceptual disparities about therapy between patient and therapist should be minded to maximize treatment benefits.

Despite the contribution of various investigators (among others, Dollard & Mowrer, 1953; Fiedler, 1953; Heine, 1953; Leary & Gill, 1959; Robbins & Wallerstein, 1959; Rogers, 1961; Strupp, 1960), agreement concerning process aspects and recognizable and predictable effects of psychotherapy is still lacking. Evaluation of psychotherapy is undoubtedly a complex and difficult area of research. Numerous approaches have been attempted (Rubinstein & Parloff, 1959; Zax & Klein, 1960), yet the authors have been impressed by the meager heed given to the patient's viewpoint of what happens in psychotherapy and its influence on him. Most reported studies have rudderless psychotherapeutic understanding and explanation mainly to the therapist's grasp.

The direct, phenomenological path (Blaine & McArthur, 1958; Board, 1959; Chance, 1959; Lipkin, 1948) has been treated with some success notwithstanding a number of obstacles, i.e., (a) neither patient nor therapist can be a completely unbiased observer, each varying in what he can and will report, (b) the mutual desire for success, wish to please the therapist, and the "hello-goodbye" effect of Hathaway (1948), may unwittingly affect responses, (c) limited consonance may exist between what either patient or therapist states about occurring changes and actual

behavior noted by others, and (d) both patient and therapist may be unaware of what facets of therapy are actually accountable for improvement and attribute success to irrelevant factors, overlooking those having meaningful and lasting effects.

Some of these impediments can be partially overcome for the patient by extending the distance between therapist and research operation in his thinking, e.g., assuring him that replies will be kept confidential, getting the questionnaire administered by and returned to someone other than the therapist, and emphasizing the impartial nature of the investigation. Additionally, follow up at a later time may yield data more liberated from immediate "transference" effects, albeit subject to possible memory distortions of their own.

Our view is, despite the stumbling blocks marring this way of advance, that the involved parties in psychotherapy are still in the most favored position to provide us with promising leads concerning what takes place. To maximize comprehension of the therapeutic experience, harmful as well as positive aspects should be examined. Undesirable features of psychotherapy have been neglected in most investigations. Avoidance of these is understandable. It appears worthwhile, however, to hypothesize that certain features of therapy may be useful and others not, and to look for evidence of both in assessing therapeutic functioning.

<sup>1</sup>A condensed version of this paper was presented at the 1962 American Psychological Association annual convention in St. Louis.

The present study was designed to serve as the exploratory phase of a research program aimed at probing helpful and nonhelpful aspects of psychoanalytically oriented psychotherapy as practiced in a Veterans Administration Mental Hygiene Clinic. The initial phase accounted for here analyzes the perceptions of both patients and therapists at the close of a mutually engaged-in psychotherapy along these lines. The therapists' abilities in predicting responses of their patients to questions about the therapy are also appraised. Additionally, stability of the patients' judgments concerning their therapy is evaluated after the passage of a number of years.

### METHOD

Choice of an instrument to measure the reactions of the participants was that of an open-end questionnaire. While other alternatives such as checklists, rating scales, *Q* sort self-descriptions, etc., can contribute useful data, the consideration was that meaningful clues might be overlooked if responses were limited to variables occurring to the investigators. Further, there was the estimation that it is difficult to fully convey exact feelings and thoughts through checking or sorting a series of statements. This does not minimize disadvantages which open-end questionnaires themselves possess, notably, (a) responses may be difficult to categorize accurately because of complexity or obscurity, (b) important variables may be ignored through oversight or pressing concern with other areas, (c) persons may assess the experience at varying levels of interpretation or inference making response comparison difficult, and (d) transposition of rich and complex data obtained in open-end inquiries into measurable categories is bound to entail a certain degree of oversimplification and distortion of meaning. Nevertheless, the judgment was that at this hypothesis seeking point the open-end questionnaire technique could serve heuristic purposes in extending insights about what happens in therapy.

Open-end questionnaires were administered to 86 veterans who had terminated a course of individual psychotherapy involving at least 10 interviews in a Veterans Administration Mental Hygiene Clinic during 1957-58. They represented the vast majority of patients completing therapy in the clinic during this time period irrespective of number of sessions. The criterion of a minimum of 10 sessions was set up to maximize meaningful replies to the questions posed. The questionnaire asked:

1. What changes, if any, do you feel occurred in you because of your therapy?
2. What about the therapy itself did you find helpful?

3. What about the therapy itself did you find not helpful, or even set you back?
4. Have you any suggestions for ways in which your therapy might have been made more helpful?

At the same time similar questionnaires were given to the 28 therapists who had these patients in therapy. The therapists included 12 psychiatrists, 12 clinical psychologists, and 4 psychiatric social workers functioning in a predominantly psychoanalytically oriented manner. They were asked to answer the first three questions and, additionally, to predict how their patients would answer these.

Sixty-three (73%) of the patients completed their questionnaires, while therapist responses were obtained from all 28 (100%). Background information concerning age, education, occupation, diagnosis, reason for termination, and a number of other factors about the patients were obtained from the clinic files and therapists.

An effort was made to reckon the extent to which the 73% return on patient questionnaires might reflect a biased sample affecting the results. Fortunately, therapist responses as well as case-history data were available on the 23 patients who had not returned their questionnaires. Replies of the therapists concerning amount of improvement judged to have occurred, types of change reported, number of therapy sessions, diagnosis, reason for termination of therapy, and factors such as age and sex were compared for those patients who had returned their questionnaires with those who had not. No significant differences were found between the two groups and it was tentatively concluded that the 73% return sample of patients was not biased in any known way.

Approximately 4 years after the original questionnaire had been sent out, a follow-up letter was sent to 61 of the original 63 participating patients (2 patients, in this interim, had returned to psychotherapy with their original therapists and it was deemed advisable to exclude them) requesting they again answer the questionnaire on the basis of how they now felt about their past therapy. Forty-five (74%) of the patients completed this follow-up questionnaire. This is a substantial feedback considering that 1 patient had died and seven questionnaires were returned because the patients could not be located. This means that only 8 patients, who apparently received their questionnaires, did not contribute to the follow-up study. Here, too, no reliable differences were in evidence with regard to diagnosis, number of sessions, previous report of satisfaction by patient, therapist judgment of degree of improvement, etc., between the 45 patients responding to the follow-up questionnaire and the 16 who did not.

### Subjects

Of the 63 patients comprising the basic sample of this study, 58 were male and 5 female. Their ages ranged from 24 to 57 with a mean of 37.5 years.



The number of years of schooling varied from 4 to 18 with a mean of 14.0 years indicating the general educational level of the group to be considerably above that of the general population of veterans. Forty percent of the group were college graduates. Occupationwise they were less outstanding, with 43% either unemployed or working at unskilled or semiskilled trades. Ten percent were diagnosed as psychotic, 36% as psychoneurotic, 35% as character disorder, and 19% as psychosomatic or psychophysiological reaction. In 78% of the cases therapy had been terminated by mutual agreement between patient and therapist, in 9% the therapist had advised termination, while in 13% the initiative had been taken by the patient. Seventy-one percent were married and an additional 13% had been married at some time.

Concerning frequency of therapy sessions, 67% had been seen on a weekly basis, 24% more often, and 9% less often. The number of cumulative interviews for the patients varied from 10 to 324, with 42% having been seen from 10 to 50 sessions, 21% from 51 to 100 sessions, and 37% for more than 100 sessions. Forty-eight percent were seen by psychiatrists, 38% by psychologists, and 14% by psychiatric social workers. About two-thirds of the patients were treated by regular staff members with a minimum of 4 years' therapeutic experience. The other third was seen by either psychiatric residents or fourth-year clinical psychology trainees. Approximately two-thirds were seen by male therapists and the rest by female therapists.

### *Classification of Responses*

Responses to the questions by both patients and therapists were sorted into categories based on analysis of their content. Those dealing with changes were adequately classified into 19 rubrics, those with helpful factors into 18 rubrics, and those covering nonhelpful aspects into 17 rubrics.<sup>2</sup> It was observed, nevertheless, that a number of the rubrics culled out still tended to overlap and be somewhat correlative. Although face reading suggested further combinations, we decided to keep these separate because of their individual citation. However, in order to obtain better comparative perspective, the response rubrics were further organized into major area categories.

Reliability of categorization was checked by having another investigator make an independent classification of a random third of the responses, using the same rubrics. Since the questions were open-end with no limit to the number of changes, etc., which a person might report, percentage of agreement was based on the total number of classifi-

cations assigned. The two investigators assigned identical category numbers to 79% of the responses. In view of the complex and often murky nature of some of the replies, this was gauged to represent an acceptable level of reliability of categorization.

## RESULTS

### *Changes Occurring in Therapy*

In response to the first question "What changes, if any, do you feel occurred in you because of your therapy?" 90% of the patients mention one or more changes for the better. The therapists of these patients also report considerable change, 84% indicating more than one beneficial change. Data on the therapists' determinations concerning the condition of these patients at the close of therapy recorded at a different time for other purposes showed that 62% of the patients were judged moderately or much improved, 30% slightly improved, and 8% unimproved or worse. Even keeping in mind the probable bias toward proclaiming favorable results, these figures imply that both patients and therapists felt that therapy was generally helpful.

Changes detailed most often by the patients are "more realistic evaluation of self" (52%), "more self-confidence" (35%), and "more ability to handle problems" (17%). Therapists of these patients list "more realistic evaluation of self" (39%), "reduction in neurotic fears" (38%), and "more self-confidence" (35%), as the most common changes. Significant differences in outlook between the groups appear concerning "improved social relationships" (chi square 10.42) and "reduction in neurotic fears" (chi square 16.58), with the therapists mentioning these changes more frequently.

When the various rubric changes are grouped into major areas (Table 1), those involving insight, followed closely by attitude and behavioral alterations, lead the patients' list, whereas those relating to behavioral and symptom relief head the therapists' entries. Both behavioral and symptom relief changes are mentioned significantly more often by the therapists than the patients. They, seemingly, note changes mirroring improved social relations and reduction in self-defeating behaviors more than the patients who emphasize intrapsychic or sub-

<sup>2</sup> These tables have been deposited with the American Documentation Institute. Order Document No. 7455 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.



TABLE 1

RELATIVE FREQUENCY OF MAJOR AREA CHANGES OCCURRING IN THERAPY

Area of change	Patients		Therapists	
	%	Rank	%	Rank
1. Insight	52	1	37	4
2. Attitude	48	2	38	3
3. Behavioral	47*	3	71*	1
4. Symptom relief	25**	4	57**	2
5. Other	10	5	8	5

Note.—Percentages represent percentage of cases giving a response classifiable in this general area.

\*  $p < .05$ .

\*\*  $p < .01$ .

jective feeling changes. One should be alert to the possibility that these may reflect dissimilarities in conceptualization rather than sharp dissonances over actual changes taking place.

### Helpful Aspects of Therapy

In answer to the question "What about the therapy itself did you find helpful?" the "opportunity to talk over problems with someone" is placed high by both patients (21%) and therapists (19%) as is "catharsis; outlet for tensions, help in expressing feelings," with 19% in each group specifying this aspect of therapy as advantageous. Both groups also stress "attitudes and characteristics of the therapist" (16%) as a positive therapeutic feature. The therapists, additionally, highlight "support especially in time of crisis" (19%) and the "transference relationships" (16%). These latter two items are not verbalized at all by the patients and significant differences exist between them and the therapists (chi square 11.14; chi square 8.80).

When the helpful aspects are classified into major areas (Table 2), those of talking over one's problems and the therapeutic relationship (therapist as a person) head the replies of the patients. The therapists give precedence to therapeutic skills, opportunity to talk over problems, and emotional support of the patient. Significant differences are evident in the frequency with which therapists regard therapeutic technique and satisfaction of emotional needs when compared with the patients' designation of

these facets. The therapists' cognizance of technique is understandable. It may also be postulated that difficulties which patients have in recognizing and coping with their emotional wants (e.g., gratification of dependency needs) make them less discerning of such elements as being beneficial for them. Again, the issue arises whether these observed differences characterize variance in reference points more than opposing views.

### Nonhelpful Aspects of Therapy

In response to the question "What about the therapy itself did you find not helpful, or even set you back?" over one-third (37%) of the patients state they cannot think of anything which interfered with treatment. Among those particularized, the most frequently reported statements are "therapists' feelings of irritation, anger, boredom, etc." (11%) and the "frequent change of therapist" (10%) to which they were subject. On the therapists' side, 16% assert that "nothing" was detrimental to therapy. Their sentiments about nonhelpful aspects underline mistiming in "interpretation and confrontation" (15%), termination problems (14%), and communication to the patient of negative personal feelings (14%). The only reliable difference between patients and therapists is the more frequent declaration by the patients that "nothing" was nonhelpful about their therapy.

Table 3, major nonhelpful areas of therapy, reflects essentially these findings. Therapist

TABLE 2  
RELATIVE FREQUENCY OF MAJOR HELPFUL AREAS OF THERAPY

Areas of help	Patients		Therapists	
	%	Rank	%	Rank
1. Opportunity to talk over problems	37	1	37	2
2. Therapeutic relationship (therapist as a person)	25	2	30	4
3. Therapeutic skill and technique	16**	3	38**	1
4. Administrative factors	14	4	10	5
5. Satisfaction of emotional needs (support)	6**	5	35**	3
6. Other	2	6	0	6

Note.—Percentages represent percentage of cases giving a response classifiable in this general area.

\*\*  $p < .01$ .

TABLE 3  
RELATIVE FREQUENCY OF MAJOR  
NONHELPFUL AREAS OF THERAPY

Nonhelpful area	Patients		Therapists	
	%	Rank	%	Rank
1. None	37*	1	16*	3
2. Therapeutic relationship (therapist as a person)	14	2	11	4
3. Administrative factors	13	3	21	2
4. Therapeutic skill and technique	9**	4	34**	1
5. Other	5	5	8	5
6. Absence of emotional support	2	6	5	6

Note.—Percentages represent percentage of cases giving a response classifiable in this general area.  
\*  $p < .05$ .  
\*\*  $p < .01$ .

statements significantly point up damaging features related to professional functioning when compared with those of their patients. Also, therapists express significantly fewer

“nothing was not helpful about therapy” statements than do the patients. A word of explanation may be in order to clarify the number two ranking (21%) of administrative factors by the therapists. This incorporates instances where therapy was terminated too soon not due to misjudgment on the part of the therapist but rather because of his leaving the clinic, the patient’s having to move to another town because of occupational reasons, etc.

Suggestions

Question 4 “Have you any suggestions for ways in which your therapy might have been made more helpful?” was asked only of the patients. Thirty-five percent had none to make. Of those volunteered, the principal ones are less frequent change of therapist (10%), more therapy time via longer inter-view sessions or more frequent visits (10%),

TABLE 4  
ACCURACY OF THERAPISTS’ PREDICTIONS OF MAJOR AREAS

Areas	Number of patients giving this response	Number of correct pre-dictions by therapists	% correct predictions by therapists
Changes			
Insight	33	9	27
Attitude	30	17	57
Behavioral	30	18	60
Symptom relief	16	10	63
Other	6	2	33
Helpful			
Opportunity to talk over problems	23	15	65
Administrative factors	19	3	16
Therapeutic relationship (therapist as a person)	16	8	50
Therapeutic skill and technique	10	2	20
Satisfaction of emotional needs (support)	4	0	0
Other	3	0	0
Nonhelpful			
None	23	5	22
Administrative factors	9	2	22
Therapeutic relationship (therapist as a person)	9	1	11
Therapeutic skill and technique	6	1	17
Other	3	0	0
Absence of emotional support	1	0	0

Note.—Percentages represent percentage of correct predictions based on the number of responses in this category given by patients.

and more guidance and direction from the therapist (10%). Another 10% desire "more patience, interest, and warmth" from the therapist.

### *Predictions of Therapists*

Table 4 presents the accuracy of the therapists in predicting how their patients would respond to the first three questions posed them. In the area of changes, their auspiciousness are best for symptom relief, followed closely by those related to behavioral and attitude changes. These areas, incidentally, are the major ones in which the therapists felt therapeutic changes had actually occurred.

Concerning acknowledgments by the patients of what was helpful in therapy, predictions of the therapists are most satisfactory in the areas of opportunity to talk over problems and the therapeutic relationship.

Considering the open-end nature of the questionnaire, the overall degree of prediction by the therapists is moderately respectable.

### *Background Variables*

Responses of the patients were examined pertaining to a number of background variables. These encompassed the patient's age, educational level, and occupation. Differences were in evidence only as often as would be predicted on the basis of chance. Further assessment was done of the kind of therapy attempted (intensive, supportive) and diagnosis. Again, reliable differences in responses were found only as frequently as would be anticipated by chance. Additional analyses of the answers were made in terms of professional discipline, sex, and experience level of the therapist. Here, too, no important relationships emerged. It made little difference whether the patient had seen a social worker, psychiatrist, or psychologist, had a male or female therapist, or been in therapy with a staff member, advanced level psychology trainee, or psychiatric resident.

### *Follow-up Findings*

As noted previously, 4 years after their therapy had ended, 45 (74%) of the patients again answered the questions originally asked

of them. The aim was to gain information about the patients' views of their therapy after they had been graduated, so to speak, in the outside world—with transference feelings presumably muted and perspectives heightened. Responses to the questions were coded afresh using the identical rubrics as before.

Seventy-eight percent of the patients now mention one or more positive changes compared with 90% previously. The specific changes itemized most repeatedly are still essentially the same as originally asked of them. "More realistic evaluation of self" (42%) heads the list, followed by "more ability to handle problems" (25%), and "more self-confidence" (22%).

The major area of change reported now (Table 5) by the patients is behavioral (47%) whereas before it was insight (52%). However, their follow-up responses are not significantly different than their original ones.

With respect to the question "What about the therapy itself did you find helpful?" the patients again highlight the same aspects they did initially. The major ones mentioned are the "opportunity to talk over problems with someone" (22%), "catharsis; outlet for tensions, help in expressing feelings" (20%), and "attitudes and characteristics of the therapist" (14%). The status quo holds concerning major areas of help (Table 5) with opportunity to talk over one's problems and the therapeutic relationship remaining the leaders. Although not reliable, there is now a lessening tendency to select administrative factors as being helpful.

Follow-up responses to the third question "What about the therapy itself did you find not helpful or even set you back?" do signal differences from the patients' original replies. The most frequent nonhelpful aspects reported now are "therapist too passive, permissive" (22%) and "interpretations and confrontations given when not prepared for them" (20%). There is a noticeable fall-off in the "nothing, everything was helpful" rubric from an original 37% to 20% (Table 5). Examination of the nonhelpful areas shows that faultfinding is centered mainly on the therapist's professional skills. This critical orientation is presently significant at



TABLE 5  
COMPARISON OF THERAPIST AND ORIGINAL PATIENT MAJOR AREA  
RESPONSES WITH FOLLOW-UP RESPONSES OF PATIENTS

	Therapists		Patients (original)		Patients (follow-up)		$\chi^2$ Therapists versus patients (original)	$\chi^2$ Therapists versus patients (follow up)	$\chi^2$ Patients (original) versus patients (follow up)
	f N = 63 <sup>a</sup>	%	f N = 63	%	f N = 45	%			
<b>Changes</b>									
Insight	23	37	33	52	17	38	2.60	0.00	1.70
Attitude	24	38	30	48	14	31	.81	.30	2.32
Behavioral	45	71	30	47	21	47	6.46*	5.77*	.01
Symptom relief	36	57	16	25	13	29	11.82**	7.35**	.03
Other	5	8	6	10	8	18	0.00	1.56	.94
<b>Helpful</b>									
Opportunity to talk over problems	23	37	23	37	14	31	0.00	.14	.14
Therapeutic relationship	19	30	16	25	8	18	0.16	1.54	.50
Therapeutic skill	24	38	10	16	5	11	6.81**	8.41**	.18
Administrative factors	6	10	9	14	1	2	.30	—	3.22
Satisfaction of emotional needs	22	35	4	6	5	11	14.01**	6.72**	.28
Other	0	0	1	2	4	9	—	—	—
<b>Nonhelpful</b>									
None	10	16	23	37	9	20	5.91	.09	2.68
Therapeutic relationship	7	11	9	14	6	14	.07	.02	.02
Administrative factors	13	21	8	13	10	22	.91	0.00	1.10
Therapeutic skill	21	34	6	9	20	44	9.24**	.94	15.65**
Other	5	8	3	5	6	13	—	.35	—
Absence of emotional support	3	5	1	2	1	2	—	—	—

<sup>a</sup> Although the N of therapists involved is 28, they responded individually for each of the 63 patients originally involved. Hence, our N reflects 63 therapist-patient pairs.

\*  $p < .05$ .

\*\*  $p < .01$ .

the .01 level when contrasted with the patients' original replies. The patients, at this time, vent displeasure and specify discontent more openly and clearly than heretofore, 95% now doing so compared to 43% previously (Table 5). Administrative factors are ranked as the second most nonhelpful area.

Concerning the final question "Have you any suggestions for ways in which your therapy might have been made more helpful?" the new suggestions indicate no reliable differences from those mentioned 4 years back. But, as manifest in responding to the nonhelpful question, there is enhanced "speaking-up." Only 13% of the patients now fail to make suggestions as contrasted with 35% originally. Among the most persistent ones made are those related to technique, i.e., "more direction and active participation from the therapist" (10%), "in-

creased length and frequency of interviews" (11%), use of hypnosis, joint therapy sessions with spouse, etc. Assuming larger proportion now are recommendations connected with administrative features, e.g., less frequent change of therapist, more attractive rooms, a shorter waiting period for therapy, more medical examinations, etc. Additional warmth and personal interest from the therapist continue to be called for.

We thought it worthwhile to also compare the follow-up responses of the patients with those of the therapists and see if any directional shifts had taken place. In the matter of changes, the therapists maintained their significant noting of behavioral and symptom relief changes when compared with the patients. In helpful areas, the therapists still disclose reliable accentuation of therapeutic skill and satisfaction of emotional needs in contrast to the patients. In the nonhelpful

realm, the patients have come around to accepting the therapists' concern with technique. The previous significant difference between the two groups on this score has vanished. Indeed, percentage-wise, this area is mentioned more often by the patients than the therapists. Additionally, the past statistical difference in the area of "nothing was nonhelpful" has disappeared in the wake of the patients' sharpened self expression.

Generally, the effect of 4 years intervening time on these patients has been to clarify what they did not like about therapy while maintaining their impression of positive benefit.

#### DISCUSSION

A thought-provoking contrast is in the patients' accent on insight changes compared with those of symptom relief and behavior by the therapists. This seems to limn the touchstone by which each group tests the world. The patient's is his inner subjective feelings; the therapist's his patient's freedom from shackling behaviors and acquired social accommodation.

Patients, in substantial proportion, attribute their assistance to being able to talk with someone about their difficulties in an atmosphere of interest, warmth, and tolerance. In opposing vein, a goodly number of the therapists conceive their aid as issuing strongly from their professional mastery. Undoubtedly idols of the den are at work here. But, frames of reference aside, hierarchical precedence, if not actual weight differences in the therapeutic multiple regression equation, are manifest. The leitmotiv that sounds again and again in the patients' replies is the importance of sharing uncertainties and urgencies with an individual who will listen with respect and treat with dignity their person. The administrative aspects, both helpful and nonhelpful, which are cited are mainly in the service of undergirding and firming the therapeutic relationship but zeroes in on what he is doing performance-wise. Allowing the patient to express himself, sustaining him in his emotional exigencies, recognizing the influence of attitudinal components—all are heeded by the therapist, but the kingpin remains his technical excellence.

The consideration is not that expectations and concepts of therapeutic implementation vary or are noncomplementary between therapist and patient; rather, that lack of awareness of these diverging calculations and hopes may often make for avoidable disruption and lag, particularly in the early stages of therapy. Interestingly enough, after the patients obtain distance from their therapy experience, they tend, although not significantly, to minimize insight and attitude changes. They veer sharply toward skill and proficiency areas in discussing what was not salutary to their therapy. This may well echo the patient's newly won capacity to attend to the beat of sociocentric drums. Yet, one wonders about the relation of this fresh verbalization of what was presumably undifferentiated or not understood before to befitting consensual "explanatory" knowledge.

Be that as it may, an implication of force is that the therapist's contribution, as far as the patient observes it, is personal as well as technical, a theme which Carl Rogers has been justly heralding for some time. Of pertinence is not only the personal feelings (countertransference) of the therapist as affirmed by Freud, Sullivan, and others but also the patient's perception of them. The notion that the therapist "cures" with what he himself is as well as his studied arts must be troika-ed with the patient's anticipations and reactions to these. It is plain that cognitive characteristics of treatment are only one of its meaningful components. Our findings, leastwise from the patients' outlook, tend to bolster Rogers' view that attitudinal elements in the therapy relationships are a consequential ingredient accounting for change. Whether they are more salient than technique is not distinct but they apparently possess subjective antecedence for a considerable number of the patients.

The enlarged capacity of the patients, at long last, to forthrightly speak up concerning "lacks" in their previous therapy lends credence to the idea that "transference" feelings are far from dissolved in many patients at the termination of their therapy. Pfeffer's (1959) work certainly reinforces this datum.

Our findings are congenial with those of



Board (1959). He also studied patient and physician judgments concerning outcome of psychotherapy in an outpatient clinic. His results showed that patients equated successful therapy with being permitted to express their problems, gaining self-insights, and having interested and understanding therapists. Unsuccessful therapy was associated essentially with noninterest on the part of the therapist and assignment to multiple therapists. This aspect of frequent change of therapist is an accompanying one in our data. It behooves clinic administrators to temper this development as far as feasible. It is humbling to realize the rejection and "lost" feeling which so many patients experience on being transferred from one therapist to another, particularly when it is not linked to therapeutic motive—this despite preparation by the therapist.

A few words of caution. We should not overlook the variety of aspects which both patients and therapists relate are helpful and nonhelpful. For instance, note that a small but definite number of patients feel that "guidance and direction" are useful to them. Some are neurotic. A psychotic patient voices the opinion that his therapist was "too controlling" for him. We must guard against the thinking that "since this is the kind of thing most patients indicated was helpful, therefore it is the thing most helpful in therapy." Clinical experience has demonstrated steadily that the same type of therapy is not necessarily most helpful to all patients even within identical nosological groups. We need to understand and integrate the information and implications for therapy offered by both our majority and minority groups among patients and therapists.

Another caveat. These are data from patients responding to a prevailing psychoanalytically oriented form of psychotherapy. Would they be similar in reacting to another theoretical style of psychotherapy?

Finally, it is somewhat reassuring to obtain indications that positive changes do take place as a result of therapy. This, notwithstanding the tendency for patients and therapists to highlight its favorable outcomes.

## REFERENCES

- BLAINE, G. B., & McARTHUR, C. C. What happened in therapy as seen by the patient and his psychiatrist. *J. nerv. ment. Dis.*, 1958, 127, 344-350.
- BOARD, F. A. Patients' and physicians' judgments of outcome of psychotherapy in an outpatient clinic. *AMA Arch. gen. Psychiat.*, 1959, 1, 185-196.
- CHANCE, ERIKA. *Families in treatment: From the viewpoint of the patient, the clinician and the researcher*. New York: Basic Books, 1959.
- DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald Press, 1953.
- FIEDLER, F. E. Quantitative studies on the role of therapists' feelings toward their patients. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald Press, 1953.
- HATHAWAY, S. R. Some considerations relative to nondirective therapy. *J. clin. Psychol.*, 1948, 4, 226-231.
- HEINE, R. W. A comparison of patients' reports on psychotherapeutic experience with psychoanalytic, nondirective, and Adlerian therapists. *Amer. J. Psychother.*, 1953, 7, 16-23.
- LEARY, T., & GILL, M. The dimensions and a measure of the process of psychotherapy: A system for the analysis of the content of clinical evaluations and patient-therapist verbalizations. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.
- LIPKIN, S. The client evaluates nondirective psychotherapy. *J. consult. Psychol.*, 1948, 12, 137-146.
- PFEFFER, A. Z. A procedure for evaluating the results of psychoanalysis. *J. Amer. Psychoanal. Ass.*, 1959, 7, 418-444.
- ROBBINS, L. L., & WALLERSTEIN, R. S. The research strategy and tactics of the psychotherapy research project of the Menninger Foundation and the problem of controls. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.
- ROGERS, C. R. *On becoming a person: A therapist's view of psychotherapy*. Boston: Houghton Mifflin, 1961.
- RUBINSTEIN, E. A., & PARLOFF, M. B. (Eds.). *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.
- STRUPP, H. H. *Psychotherapists in action*. New York: Grune & Stratton, 1960.
- ZAX, N., & KLEIN, A. Measurement of personality and behavior changes following psychotherapy. *Psychol. Bull.*, 1960, 57, 435-448.

(Received October 11, 1962)



## BODY CATHEXES OF PARENTS OF NORMAL AND MALFORMED CHILDREN FOR PROGENY AND SELF

LOUISE CENTERS

AND

RICHARD CENTERS

*St. John's Hospital, Santa Monica*

*University of California, Los Angeles*

As a test of the hypothesis that malformation in the body of the child constitutes a threat to the bodily integrity of the parent, and will result in body-cathexis responses differing from those of parents of normal children, two forms of the Body-Cathexis Scale were administered to samples of parents of both malformed and normal children. Support for the hypothesis was discovered in that there was a substantial correlation between responses to their children's bodies and responses to their own bodies on the part of parents of the malformed children, while little or no correlation existed in this respect for the parents of normal children. Parents of malformed children also expressed more dissatisfaction with their children's upper extremities.

Freud and Adler early in their careers manifested an awareness that the individual's reactions to his body constituted some of the most interesting of psychological phenomena, calling attention to and writing extensive commentaries on such aspects of the body-self relationship as anxiety, narcissism, and the inferiority complex. Curiously, psychologists as well as laymen have seemed to accept what insights these pioneers had to offer with little attempt at empirical check on their validity. Schilder's (1950) effort to elicit beliefs, feelings, and memories about the body in questionnaires developed by him as adjuncts to group and individual therapy represent the first attempts at objective and systematic collection of data. But by far the most serious and ambitious strivings in this direction are found in the work of Jourard and Secord (1954, 1955a, 1955b; Secord, 1953; Secord & Jourard, 1953). They devised a set of questions to form a more objective scale than Schilder had provided, terming their instrument a Body-Cathexis (B-C) Scale. Secord and Jourard (1953) wrote: "By *body-cathexis* is meant the degree of feeling of satisfaction or dissatisfaction with the various parts or processes of the body [p. 343]." This is, of course, a somewhat different use of the term than the psychoanalytic meaning, which indicates the amount of instinctual energy invested in one's body as an object.

The B-C Scale, which is described more

fully later, consists essentially in requiring the subject to indicate his degree of satisfaction or dissatisfaction with each of various parts of his body in terms of a five-point scale. The procedure, although subject to some of the objections and criticisms common to questionnaires and self-inventories, has the advantages of relative simplicity, proven value in a substantial amount of previous work, and demonstrated internal consistency and reliability (a reliability coefficient of .81 has been obtained with it).

Of most relevance to the research reported here is a study by Remy (1953). He used college males and females as subjects and gave them both body-cathexis (B-C) and self-cathexis (S-C) questionnaires. All subjects answered three questionnaires on each variable, filling out one for representing their own feelings on each, another two as they believed their mothers felt about their (the subjects') bodies and selves, and a third pair representing their perception of their fathers' feelings. All subjects also took Maslow's Test of Psychological Security-Insecurity (Maslow, Hirsh, Stein, & Honigman, 1945).

The main results of the study were that *perceived* mother and father attitudes about the subjects' bodies were positively correlated to a substantial ( $r$ 's ranged from .56 to .84) and significant degree with self-evaluations, and that both parental cathexes and BC-SC scores correlated to a substantial degree with feelings of security.

### Problem

If the foregoing correlates of body cathexis seem somewhat a demonstration of the obvious, they are no less valuable for that. They are specially worth while, moreover, in suggesting validity for the scale itself.

A distinctly less obvious correlate or determinant of body cathexis in the individual may lie in the social evaluation of the body of his progeny by himself and others. In instances where from birth or later accident the body of the child is in some way malformed, rejection and subtle defensive maneuvers may be found, though the existing evidence which would bear upon the foregoing point is at most quite general and indirect.

For instance, Wittreich and Radcliffe (1955) had subjects wear aniseikonic lenses of varying strengths when viewing two figures, one normal, the other mutilated. On a scale purporting to measure "resistance to induced aniseikonic distortion they found that the threshold for the mutilated figure was significantly higher than for the normal figure." The investigators in this case decline to offer any plausible interpretation of their findings, but to us a reasonable suggestion is that the phenomenon is a manifestation of perceptual defense.

Gilder, Thompson, Slack, and Radcliffe (1954) achieved similar results to those of the foregoing investigation when they utilized full-sized mannequins as well as adult amputees as stimuli. One of their findings was that the perception of the amputee and even the mutilated mannequin was apt to provoke strong emotions of anxiety as well as anger. They interpret these reactions as a function of the threat that actual or even simulated amputation constitutes to the integrity of the body image of a normal person since he identifies with the person viewed. The type of emotional response encountered is seemingly dependent upon the meaning the body injury has for the subject and his particular way of dealing with such stress.

If persons unrelated to amputees experience such threat and manifest such emotional reactions, then how much more likely it would be that the parent of an amputee

child will experience threat and manifest defensiveness in anger and rejection? The question of whether he does or not is significant from several considerations, the most obvious of which is that of the psychological adjustment of both parent and child to the circumstance of the latter's deformity. Feelings of shame, guilt, and resentment have all been observed in parents of amputees by the senior author in clinical interviews with such parents in connection with her work with the Child Amputee Research Project at the University of California, Los Angeles, although how such reactions are related to bodily integrity and body cathexis was not obvious. That the family stresses imposed by chronic *parental* illness are accompanied by affective and defensive reactions on the part of the children of such parents has been effectively shown in Arnaud's (1959) study. That effects may be produced in an opposite direction seems highly probable.

Out of the beliefs that the amputee child does constitute a threat to the bodily integrity of the parent, it was expected that reactions to such threat might be manifested in the responses to the Body-Cathexis questionnaire. More specifically it was our hypotheses that (a) parents would respond with rejection of those parts of their child's bodies where malformation of the child's body existed, and thus would show extreme dissatisfaction for them in terms of the B-C scale, and that (b) because of identification with the child, a substantial correlation in the responses of parents of amputee children toward their own and their children's bodies would be found.

### PROCEDURE

For testing the foregoing hypotheses two samples of parents, one possessing a malformed child and the other a normal one, were employed as subjects. Both children and parents in the respective samples were matched in as many ways as possible to insure that whatever effect parentage of a malformed child might have on response to the B-C test would not be confounded with the effects of other possible determinants.

Twenty-six parents, 6 men and 20 women whose children were victims of upper extremity amputations or anomalies, were selected for the "experimental" group. Their children were in the age range from 5 to 12 years and in all but four cases had



congenital amputations or anomalies, while the remaining four cases had been victims of disease or accidents before they were 6 years old. The children were otherwise in good general health, possessed normal vision without correction, and were of at least average intelligence as measured by the Columbia Test of Mental Maturity (Burgermeister, Blum, & Lorge, 1954). Each of these children was matched with a normal child in chronological age, mental age, and sex, and the parents of the latter were chosen as a "control" sample. Five of the parents selected for the latter group, who had initially agreed to participate, later declined to cooperate in the study, thus reducing the number in the final sample to 21. As finally constituted it consisted of 18 women and 3 men. The two parent groups were to a high degree alike in most ascertainable characteristics.

One parent of each child with an amputation and one parent of a child without one were asked to complete two forms of the Body-Cathexis Scale, one indicating their attitude toward their child's body and the other their feelings about their own bodies. The scale for child's body was given first. The version of the B-C scale employed consisted of a listing of 46 parts or aspects of the body which were to be rated by subjects according to the following 5-point scale: 1—strongly dislike and wish change could somehow be made; 2—don't like, but can put up with; 3—have no particular feelings one way or the other; 4—am satisfied; 5—consider myself fortunate. Thus, if a person were very dissatisfied with his nose, for example, he would encircle a 1.

The 46 body parts or aspects are: hair, facial complexion, appetite, hands, distribution of hair over body, nose, fingers, elimination, wrists, breathing, waist, energy level, back, ears, chin, exercise, ankles, neck, shape of head, body build, profile, height, age, width of shoulders, arms, chest, eyes, digestion, hips, skin texture, lips, legs, teeth, forehead, feet, sleep, voice, health, sex activities, knees, posture, face, weight, sex (male or female), back view of head, trunk.

In keeping with Johnson's recommendation for reducing response sets the following paragraph was

added to the original instructions: "You may refer back to the scale as often as necessary to make your judgment of how you feel. Judge each item carefully. Do not use the same number for each item."

### RESULTS AND DISCUSSION

Frequency distributions of the total scores and of each of the five ratings on the B-C scale, Child form, were made for the amputee and nonamputee groups. No significant differences were found between the two with regard to their median scores.

Since only four items in the scale pertained to the upper extremities, it was thought possible that a differential treatment of these four items by the parents of amputees might be masked by the distribution of ratings of the remaining items. For this reason responses to these four items, Hands, Fingers, Wrists, and Arms, were extracted from the scale, and subjected to separate chi square tests. Because of the small cell frequencies Rating Categories 1 and 2 were combined, as was 4 with 5. The results of these tests are presented in Table 1. They indicate strong dislike for those features of the amputee child's body which are deformed, such as fingers, hands, wrists, and arms.

No significant differences between the groups were found, however, on the remaining items. In other words, the parents of amputee children strongly dislike those features of their children's bodies which are deformed, but this attitude does not extend to other parts or functions of the child's body nor to the body as a whole. The results of this test, then, tend to support the hypothesis that parents of amputee children will have

TABLE 1  
CHI SQUARE TESTS OF RATINGS FOR ITEMS PERTAINING TO UPPER EXTREMITIES: CHILD FORM

Rating	Hands		Wrists		Fingers		Arms	
	Amputee	Non-amputee	Amputee	Non-amputee	Amputee	Non-amputee	Amputee	Non-amputee
1 + 2	9	0	3	0	6	0	9	0
3	4	5	5	4	4	3	2	2
4 + 5	13	16	18	17	16	18	15	19
	$\chi^2 = 9.2^*$		$\chi^2 = 2.23$		$\chi^2 = 5.7$		$\chi^2 = 12.96^*$	

\*  $p < .01$ ; Yates correction for small frequencies.



TABLE 2  
CORRELATIONS BETWEEN PARENT'S ATTITUDE  
TOWARD CHILD'S BODY AND ATTITUDES  
TOWARD OWN BODY

Scale	Amputee Parent-Child (N = 26)	Nonamputee Parent-Child (N = 21)
46-item	.49*	.09
42-item	.54*	.07

\* $p < .01$ .

extreme attitudes of dislike toward their children's bodies insofar as they reject the specific anomalous feature of the child.

The prediction was also made that on the Body-Cathexis Scale, Parent form, there would be a substantial correlation between the attitudes the parents of amputee children expressed toward their children's bodies and those they expressed about their own bodies. To test this prediction an analysis of the Parent form similar to that of the Child form was made. No significant differences were found between the groups of parents, either in their attitudes toward their bodies, all parts and functions, or toward their own upper extremities. In effect, parents of amputee children are apparently just as satisfied with and accepting of their own bodies as are parents of nonamputee children.

The extent to which parents' attitudes toward their children's bodies agree with attitudes toward their own bodies was ascertained by computing Spearman rank correlations. In each case the total score for responses to the Child form was paired with scores on the Parent form. Correlations were computed for total scores for both the 46-item scale and the 42-item scale and are shown in Table 2. It is apparent that there are substantial correlations (.49 and .54) between the attitudes expressed toward the bodies of amputee children by their parents and the attitudes they express toward their own bodies. However, there is very little relationship ( $r = .09$  and  $.07$ ) between the attitudes of parents of nonamputee children toward their children's bodies and toward their own bodies.

Although the hypothesis is supported that

the parents of amputee children have attitudes toward their own bodies similar to those they have toward their children, it appears, from scrutiny of responses to individual items, that they do not reject their own upper extremities as they do their children's. Rather they distribute their dissatisfactions among the various other parts and functions of their bodies.

A possible interpretation of this latter finding is that the strong emotional involvement and identification with the child on the part of the parent results in a generalized criticalness and dissatisfaction with own body parts which, while *normal*, are not *ideal*, and the net result is a body cathexis for self which bears an overall resemblance to that for the offspring.

It is strongly suggested, as a general hypothesis needing confirmation in further investigation, that for the parent of the deformed child his offspring may represent such an intimate extension of self that he is on the whole much less able to differentiate his evaluation of his child's body from evaluation of his own than is the parent of the normal child. Further, the findings strongly suggest the need for more general examination of the question of how perception of defect and deformity in others affects the individual's perception and evaluation of self.

# REFERENCES

- ARNAUD, SARAH H. Some psychological characteristics of children of multiple sclerotics. *Psychosom. Med.*, 1959, 21, 8-22.
- BURGERMEISTER, BESSIE B., BLUM, L. H., & LORGE, I. *Columbia Mental Maturity Scale manual*. New York: World Book, 1954.
- GILDER, R., THOMPSON, S. V., SLACK, C. W., & RADCLIFFE, K. B. Amputation, body image, and perceptual distortion: A preliminary study. *USN Med. Res. Inst. Rep.*, 1954, 12, 587-600.
- JOURARD, S. M., & SECORD, P. F. Body size and body-cathexis. *J. consult. Psychol.*, 1954, 18, 184.
- JOURARD, S. M., & SECORD, P. F. Body-cathexis and the ideal female figure. *J. abnorm. soc. Psychol.*, 1955, 50, 243-246. (a)
- JOURARD, S. M., & SECORD, P. F. Body cathexis and personality. *Brit. J. Psychol.*, 1955, 46, 130-138. (b)
- MASLOW, A. H., HIRSH, E., STEIN, M., & HONIGMAN, I. A clinically derived test for measuring psychological security-insecurity. *J. gen. Psychol.*, 1945, 33, 21-41.

- REMY, R. Relationships among body and self-cathexes, perceived parental cathexes and security. Unpublished master's thesis, Emory University, 1953.
- SCHILDER, P. *The image and appearance of the human body*. New York: International Universities Press, 1950.
- SECORD, P. F. Objectification of word association procedures by use of homonyms: A measure of body-cathexis. *J. Pers.*, 1953, 21, 479-495.
- SECORD, P. F., & JOURARD, S. M. The appraisal of body-cathexis: Body-cathexis and the self. *J. consult. Psychol.*, 1953, 17, 343-347.
- WITTEICH, W. J., & RADCLIFFE, K. B., JR. The influence of simulated mutilation upon the perception of the human figure. *J. abnorm. soc. Psychol.*, 1955, 51, 493-495.

(Received October 15, 1962)

# PSYCHOLOGICAL CORRELATES OF SOMATIC COMPLAINTS IN PREGNANCY AND DIFFICULTY IN CHILDBIRTH<sup>1</sup>

MARVIN ZUCKERMAN

*Adelphi College*

JOHN I. NURNBERGER, SPRAGUE H. GARDINER, JAMES M. VANDIVEER

*Indiana University Medical Center*

BEATRICE H. BARRETT

*Metropolitan State Hospital, Waltham, Massachusetts*

AND ARIE DEN BREEIJEN

*State Hospital, Hollywood, Florida*

The purpose of the study was to test the hypothesis that somatic complaints during pregnancy and difficulties in childbirth are related to anxiety and conflicts involving attitudes and feelings toward others. The Ss, 52 primiparous women, were given a series of tests and structured interviews. Significant positive relationships were found between somatic complaints during pregnancy and manifest anxiety, marital conflict, and indirect indicators of conflict in the interview, i.e., history of menstrual symptoms, birth, and baby dreams. 49 of the Ss' delivery room records were rated for duration of labor and amount of analgesic required. Only one significant relationship was found: anxiety, as measured by the Affect Adjective Check List given during pregnancy, was directly related to amount of analgesic required.

Many gynecologists and obstetricians are convinced that anxiety, stemming from conflicts about pregnancy and childbirth, is an important factor in the common somatic disturbances in pregnancy and the pain and difficulty in childbirth. These convictions have been largely supported by clinical evidence until recently.

Two studies have found responses to anxiety and psychosomatic scales related to physical and emotional symptoms during pregnancy (Zemlick & Watson, 1953; Schaefer & Manheimer, 1960). Anxiety measures from the TAT (Zemlick & Watson, 1953) and MAS (Davids, DeVault, & Talmadge, 1961) have been related to "abnormal reactions" during childbirth.

The present study represents another attempt to test the hypothesized relationships between anxiety and conflict on one side and

somatic reactions during pregnancy and difficulty during childbirth on the other. An attempt was made to specify some of the independent and dependent variables more precisely than in previous studies by using more specific measures. Anxiety and somatic complaints were assessed at monthly intervals during pregnancy rather than by depending on only one time sample. The study reported here was a pilot to a more extensive longitudinal study of pregnancy, childbirth, and childrearing being conducted in the Indiana University Medical Center.

## STUDY I

### *Procedure*

The subjects were 52 primiparous women seen at the prenatal clinic of the Indiana University Medical Center. The racial composition of this sample differs from previous studies. Forty-one of the group were Negro and 11 were white. Most subjects came from a lower or lower middle-class background. Their mean age was 21.1 (*SD* 5.2), and the mean of their years of schooling was 11.3 (*SD* 2.1).

<sup>1</sup> A note of thanks is owed to Eugene E. Levitt for his contributions in the planning stage of this study and to Dale E. Berkebile for his assistance in collecting the data.



Most of the subjects were seen during the fourth or fifth month of their pregnancy. At that time, they were given the first of a series of structured interviews designed to assess their reactions to pregnancy and significant family figures in addition to certain other medical information not relevant to this study. Descriptions of the structured interview technique will be elaborated later in this section. The subjects were also given a test battery consisting of the following scales:

1. The Taylor (1953) Manifest Anxiety Scale (MAS).

2. Zuckerman's (1960) Affect Adjective Check List (AACL), a test which uses a check list of words with affective connotations and an empirically developed anxiety key. The "Month" version of this test, used in this study, asks the subjects to check words describing their feelings during the previous month.

3. Three scales from Schaefer and Bell's (1958) Parental Attitude Research Instrument (PARI), a scale designed to measure attitudes toward child rearing and family life. The three scales used were: Marital Conflict, Rejection of the Homemaking Role, and Irritability (with children). These scales were found to comprise a "Hostility-Rejection" factor in the test (Zuckerman, Barrett-Ribback, Monashkin, & Norton, 1958).

4. The Masculinity-Femininity scale from the MMPI.

5. A form of Osgood's (1957) Semantic Differential was given to 16 of the subjects. The concepts compared were: Myself, Mother, Most Women, Ideal Woman, Father, Husband, Most Men, and Ideal Man. Difference (D) scores were computed between all of the female figure concepts and the Ideal Woman concept and between all of the male figure concepts and the Ideal Man concept. Increasing D scores between any figure and the ideal concept for that figure was taken as a measure of decreasing "acceptance" of the particular figure.

Forty-two subjects were available for a second structured interview, 1 month after the first interview. This interview contained a list of 13 somatic complaints: backache, pain in womb, headache, cramps, nausea, vomiting, constipation, diarrhea, urgency to urinate, weakness, fatigue, change in heartbeat, and difficulty in breathing. The interviewer inquired as to their presence or absence and their severity. A symptom complaint score was computed from this list by scoring 1 for any mild symptom complaint and 2 for any severe complaint. During this second session, the subject was again given the MAS and AACL.

Thirty subjects were available for a third interview containing the Somatic Complaint Inquiry along with other areas of content. They were also given the MAS and AACL again. On all subsequent monthly occasions, the subjects were interviewed, somatic complaints assessed, and MAS and AACL were readministered. The last interview was about 7 days before expected delivery. Varying numbers of interviews were obtained from different subjects

due to dropouts, varying starting times for the first interview, and termination of the study before completion of the series on all subjects. Ten subjects had one interview, 12 subjects had two interviews, 11 subjects had three interviews, and 19 subjects had four or more interviews.

Since the 10 subjects with only one interview never received a somatic complaint inquiry, they could not be used in any comparison involving this score. Since different subjects had different numbers of somatic complaint inquiries, MASs, and AACLs, average scores on these variables were calculated for each subject. A more homogeneous group consisting of 14 subjects with a minimum of five interviews and testings between the fourth and ninth months of pregnancy was analyzed separately on certain measures (Group II in Table 1).

*Structured interviews.* The structured interviews attempted to cover a wide range of topics which the authors felt might bear on psychological reactions to pregnancy. Questions ranged from very direct ones, such as "Have you been working during your pregnancy?" to very indirect or open-ended questions such as, "How did you feel when you really suspected you might be pregnant?" A number of interviewers were used ranging from medical students to psychiatrists. Interviewers were cautioned not to lead the subjects, and were asked to record responses verbatim. Some deficiencies were apparent when going over the interview transcripts. Since this was not the most verbal group of subjects, their responses were mostly bland or defensive in nature. Questions regarding feelings elicited very little. Since the interviewers had been warned against "leading" the subjects they tended to accept non-committal responses. A lack of rapport with the subjects may have also limited the material which was elicited from them. Despite the obvious deficiencies in the interview technique, an attempt was made to categorize responses in order to compare high and low symptom groups. It was hoped that leads might be obtained for devising a new interview schedule to be used in more extensive studies planned for the future.

## Results

*Marital status.* Eight pregnant subjects in the group were single at the time of the interviews. Eleven subjects had been married between 1 and 3 months. Since they were presumably in their fourth month of pregnancy, it was assumed that subjects in this group were single at time of conception. A third group of 33 subjects had been married 4 months or more. The scores of these three groups on the MAS, AACL, Somatic Complaint inquiry, M-F Scale, and three PARI scales were compared, using simple analysis of variance. Although Irritability, MAS, and AACL mean scores were somewhat higher

TABLE 1

CORRELATIONS OF MEAN SYMPTOM SCORE WITH TESTS

	Group I <sup>a</sup>	Group II <sup>b</sup>
<i>N</i>	42	14
Manifest Anxiety Scale	.34*	.59*
Affect Adjective Check List	.20	-.06
Masculinity-Femininity	.17	.05
Marital Conflict	.32*	.55*
Rejection of Homemaking Role	.10	.22
Irritability (with children)	-.19	.14

<sup>a</sup> A minimum of two testings between fourth and ninth months of pregnancy.<sup>b</sup> A minimum of five testings between fourth and ninth months of pregnancy.\*  $p < .05$ , two-tailed test.

in the "single" group, none of the  $F$  ratios reached the .05 level of significance.

*Correlations between tests and somatic scores.* Table 1 presents the correlations between the objective test scores and the somatic complaint scores obtained from the interviews. Group I contains all subjects who received at least one symptom complaint inquiry. Group II contains only those subjects with a minimum of five interviews. Group I includes the subjects in Group II. In both the extended and restricted samples, the MAS and the Marital Conflict scale from the PARI correlated significantly with the somatic symptom complaint score.

*Semantic differential measures and symptom complaints.* The correlations between the six D scores and the somatic symptom score are listed in Table 2. All of the correlations, with the exception of that one involving the Ideal Man-Father discrepancy, were positive. However, the only correlation which approached significance was that between the Ideal Woman-Mother discrepancy and the Somatic Symptom Complaint score. The tendency was for less acceptance of "Mother" to be associated with higher somatic complaints during pregnancy.

*Somatic symptom complaints and the structured interview.* The highest 11 subjects on the average somatic symptom complaint score were compared with the lowest 11 subjects on their responses to the first interview. All single subjects were excluded from the comparison since many of the questions in the interview were irrelevant to their situation. Of the 44

questions contained in the first interview, many pertained to medical history, and others obviously elicited little in the way of categorizable responses. Ten questions were selected for analysis. These were questions pertaining to feelings of the subjects and their husbands about the pregnancy, plans for method of feeding the baby, history of menstrual difficulty, source of, and ways of handling hostility, and dreams. Two questions yielded significant differences between the two groups. Direct questions on feelings about being pregnant and having a baby revealed no differences between the two groups. There was a significant difference between groups on reports of premarital menstrual difficulties (usually cramps). Seven of the 11 subjects in the high somatic symptom group reported these difficulties while only 2 of the 11 subjects in the low group reported them. A second significant difference was found on the reporting of dreams about babies and birth with 6 subjects in the high symptom complaint group and only 1 subject in the low symptom complaint group reporting such dreams.

During the second interview, the subjects received the first Somatic Symptom Complaint Inventory. The selection of high and low somatic symptom groups was based on average symptom scores. However, even at this second interview, there was no overlap between the two groups: the scores of the low group ranged from 0 to 2, and the scores of the high group from 3 to 9. The other questions in the second interview covered

TABLE 2  
MEAN D SCORES AND CORRELATIONS WITH SYMPTOM SCORES<sup>a</sup>

	Mean D	$r$ with Somatic symptoms
Ideal Woman versus Self	5.83	.30
Ideal Woman versus Most Women	7.67	.15
Ideal Woman versus Mother	6.51	.46*
Ideal Man versus Father	5.80	-.09
Ideal Man versus Husband	4.91	.34
Ideal Man versus Most Men	6.57	.38

<sup>a</sup>  $N = 16$  except in the case of the Ideal Man versus Husband where three single subjects were excluded.

\*  $p < .07$ , two-tailed test.



a wide range of topics: feelings about the pregnancy, situational stresses, past and present sexual behavior, and so on. Nineteen questions were selected for analysis, five questions yielding significant differences between groups. When asked how things had been going since their last visit, all 11 of the low symptom group said "fine" or "all right," while four of the high symptom group mentioned difficulties. One subject mentioned lack of sleep and "things" getting her "down"; another subject said she was "mixed up" and using her pregnancy to "pick arguments"; the third subject complained of gaining weight, and tiredness; and the fourth subject said "I will get sick, but I'm still living." When asked if things had been going as they expected, all 11 in the low symptom complaint group said "yes" while 4 in the high symptom complaint group said "no." When asked about changes in their sleep pattern, 10 subjects in the high symptom group and 6 in the low reported changes. About half in each group reported sleeping less while half reported sleeping more. In questions about the content of their dreams, the result found in the first interview was found again. Seven subjects in the high symptom complaint group reported dreaming about babies and birth and only 2 in the low symptom complaint group reported such dreams. More will be said about the content of these dreams in the Discussion section. The last significant result found in this interview was in regard to changes in interest since pregnancy. Seven of the high symptom complaint group reported an increased interest in babies while only one of the low symptom complaint group reported this particular interest. The interviews beyond the second were not analyzed because of the loss of subjects from the two symptom groups.

## STUDY II

### Procedure

The subjects were 49 of the 52 women who had been used in the previous study on pregnancy. The comparisons in this study utilized the psychological test data obtained in the initial testing session and the average somatic symptom complaint score obtained in the interviews.

The senior obstetrician reviewed the subjects' charts, without prior knowledge of psychological

test results, and delineated two variables: length of labor from the first contractions to the second stage of labor, and amount of analgesic required by the patient. The obstetrician believed that the latter was a rough objective index of the pain reactions of the patients during labor. The group was divided into a long labor group (15 hours or longer) of 13 subjects, an intermediate group (6-14 hours) of 24 subjects and a short labor (5 hours or less) group of 12 subjects. On the second variable, 6 subjects required no analgesic, 25 required the average amount, and 18 required above average amounts to deal with their pain.

### Results

*Length of labor.* The three duration of labor groups were compared on the two measures of anxiety (MAS and AACL) obtained during the fourth or fifth months of pregnancy, the three scores from the parental attitude scale (PARI) and the "Hostility-Rejection" factor score resulting from their combination, the Masculinity-Femininity test from the MMPI, and the average somatic symptom complaint score obtained during pregnancy. A one-way analysis of variance was computed between the three groups for each of the eight variables. None of the  $F$  values reached statistical significance at the .05 level. The only  $F$  approaching significance ( $F = 2.10$ ,  $2/46$   $df$ ) was that for the AACL. The mean AACL anxiety scores were 8.77 for the long labor, 6.78 for the intermediate, and 5.58 for the short labor groups.

*Amount of analgesia.* The three groups formed on the basis of amount of analgesic required were compared on the eight variables mentioned above. Only one of the eight  $F$  tests yielded a significant  $F$  ratio. The  $F$  between groups on the AACL anxiety measure was 4.27 which for 2 and 46 degrees of freedom is significant below the .05 level. The six subjects requiring no analgesic had a mean AACL score of 3.17, the 25 subjects in the average analgesic group had a mean score of 6.96, and the 18 subjects requiring above average amounts of analgesic had a mean score of 8.39.

### DISCUSSION

The significant correlation between the MAS and somatic complaints during pregnancy supports the idea that many of these



complaints reflect somatic expressions of anxiety, and their relationship with the Marital Conflict scale suggests the origin of this anxiety. The Marital Conflict scale is an indirect test, in that it asks the subject about marriages in general rather than about her own marriage. The more direct evaluations of the husband obtained in the Semantic Differential and the structured interview were mostly positive or neutral. In fact the D score on the Semantic Differential shows greater idealization of husband than of self, parents, or other persons. We would hypothesize that a conflict between unconscious feelings of hostility toward the husband and dependency on him may be a primary source of anxiety in pregnancy. Rejecting attitudes toward the mother were more overt and tended to be directly related to somatic symptoms.

A similar contrast exists in the interview where direct questions about the subject's own feelings toward her pregnancy and questions about worries elicited mainly positive feelings and denial of anxieties in both high and low symptom groups. Dream material, on the other hand, seemed to reflect a greater preoccupation with the forthcoming parturition in the subjects with many somatic symptoms. Some of these dreams have obvious conflictual content. For instance, one subject dreamed "that someone is taking something away from me, something about the baby." Another subject dreamed that she had a huge baby and did not know how she did it. Several subjects dreamed about having twins. One subject dreamed she had a girl, and then added that her husband wanted a boy. One low symptom subject dreamed that her baby was too small and her husband said it was not big enough to be a baby. The dream report technique seems to be of some potential value in exploring fears in pregnancy. One improvement in this technique might be to have subjects in future studies keep a pad and pencil by their night stands to record dreams immediately.

On the conscious level, the high symptom group expressed a greater interest in babies than the low symptom group. In terms of Dollard and Miller's (1950) conceptualization of conflict, given two equal avoidance

gradients, the situation with the higher approach gradient produces more conflict. Perhaps the motivations of this group are higher in regard to pregnancy, which would increase the strength of existent conflicts.

The greater tendency to report premarital menstruation difficulties in the high symptom group can be interpreted in two ways. First, one could hypothesize that earlier conflicts over sex, previously expressed in menstrual difficulties, have made this group particularly vulnerable to somatic stress involved in pregnancy. Second, one could hypothesize that this group has a general tendency to report somatic symptoms which others would forget about or ignore. The second interpretation is more parsimonious, but going back to the question on physical problems before pregnancy and analyzing responses other than menstrual difficulties, it was found that the same number of subjects (three) in both high and low symptom groups complained about other physical symptoms before pregnancy.

The results comparing the psychological test variables obtained during pregnancy with the obstetrical variables did not yield impressive findings. None of the test variables was significantly related to the duration of labor and only one of the variables, the AACL measure of anxiety, was significantly related to the amount of analgesic required by the patient during delivery. This result was in the expected direction with those subjects requiring more analgesic having higher anxiety scores earlier in pregnancy. However, the lack of significant relationships between the MAS and Attitude measures and the labor variables fails to support the findings of Zemlick and Watson (1953) and Davids et al. (1961). The differences in results may be due to differences in the populations used or different criteria used to evaluate childbirth difficulty. Future studies should devote more attention to observation of the subjects' behaviors during delivery rather than depending on delivery chart records.

#### REFERENCES

- DAVIDS, D., DEVALT, S., & TALMADGE, M. Anxiety, pregnancy, and childbirth abnormalities. *J. consult. Psychol.*, 1961, 25, 74-77.

- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- OSGOOD, C. E., SUCI, F. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.
- SCHAEFER, E. S., & BELL, R. Q. Development of a Parental Attitude Research Instrument. *Child Developm.*, 1958, 29, 339-361.
- SCHAEFER, E. S., & MANHEIMER, H. Dimensions of perinatal adjustment. Paper read at Eastern Psychological Association, New York City, April 1960.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- ZEMLICK, M. F., & WATSON, R. I. Maternal attitudes of acceptance and rejection during and after pregnancy. *Amer. J. Orthopsychiat.*, 1953, 23, 570-584.
- ZUCKERMAN, M. The development of an Affect Adjective Check List for the measurement of anxiety. *J. consult. Psychol.*, 1960, 24, 457-462.
- ZUCKERMAN, M., BARRETT-RIBBACK, BEATRICE, MONASHKIN, L., & NORTON, J. A. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, 22, 165-171.

(Received October 29, 1962)

## RESPIRATION AND GSR AS FUNCTIONS OF WHITE SOUND IN SCHIZOPHRENIA<sup>1</sup>

VLADIMIR PISHKIN<sup>2</sup>

*Behavioral Science Laboratory, Veterans Administration Hospital, Oklahoma City*

AND DAVID HERSHISER

*State University of Iowa*

The hypothesis was that white sound would reduce stress reaction. The 4 groups were: 20 normals-white sound (N-WS), 20 normals-no white sound (N-NoWS), 20 schizophrenics-white sound (S-WS) and 20 schizophrenics-no white sound (S-NoWS). GSR conditioning and respiration rates were measured for 20 trials, with a light as a CS and an electric shock as a UCS. The significant findings were: (a) S-WS had highest respiration, (b) changes in GSR conductance were greatest in N-WS, (c) basal GSR increased only in N-WS, (d) schizophrenics groups failed to condition. The conclusions were: (a) WS is a noxious stimulus (70 db.); (b) 2 groups responded with distinguishable physiological patterns; (c) fear and anger constructs were inferred for schizophrenics and normal groups, respectively; (d) lower oxygen saturation in schizophrenics was inferred in explanation of findings.

Psychogenic disorders, often manifested by tension, agitation, or anxiety, have been partially overcome by use of psychotropic drugs capable of producing toxic effects. Possible application of white sound (WS) as another therapeutic modality becomes obvious by virtue of its success as an audio analgesic in minor medical and dental surgical procedures (Cook, 1960). This study was undertaken in an attempt to evaluate the function of WS with schizophrenic and normal subjects under induced stress. The available information does not provide hints as to the nature of influence that may be exerted by WS on schizophrenic and normal populations. However, there is experimental evidence (Malmo, Shagass, & Davis, 1951) that schizophrenic patients respond to stress evoking situations with a greater degree of muscular tension than normal subjects. Whether WS would function to reduce induced anxiety can not be predicted, although there are data suggesting that it would be effective as an analgesic.

<sup>1</sup> This study was supported, in part, by the Veterans Administration Medical Research Program (8200 Funds).

Grateful acknowledgment is made to Harry F. Harlow, Primate Laboratory, University of Wisconsin, for his thoughtful suggestions with respect to this investigation.

<sup>2</sup> Also affiliated with the University of Oklahoma Medical Center.

Recently there has been some success in application of WS as an analgesic (Gardner & Licklider, 1959) in dental operations, such as cavity preparation, sealing, grinding, and extraction. They found that for 63% of 600 patients, the WS was completely effective, 25% of the patients found it less than complete, but no other anesthetic was needed, and 12% were not influenced.

The WS is much like white noise in that all frequencies are simultaneously present, but unlike white noise, the intensity of each frequency is not equal, but tailored to conform more closely to frequency sensitivity of the ear. According to Cook (1960),

White sound is derived from the very precise tailoring of white noise as a raw material to a new discipline of distribution of intensity versus frequency, which matches the sensitivity versus frequency characteristic of the ear . . . white sound is capable of producing uniform and high degrees of stimulation of a large fraction of the auditory nerve endings as they are distributed along the basilar membrane of the cochlea [p. 31].

There have been no reports of WS being used in a laboratory situation with induced stress. However, there is also evidence that noise in itself can produce stress. Numerous reports in closely related literature on the effects of noise would indicate that possibility of stressful influence can not be excluded.



Duncker (1937) found that a severe pain in humans resulting from pressure applied to one arm was reduced when the other arm was also stimulated painfully, and also the same effect was achieved when he used a loud noise in place of the second stimulus. There are other studies demonstrating the noxiousness of loud noise. Some of the results that have been obtained are: an increase in trembling resulting from a 90-db., 8,000-cps tone (Miller, 1953); impaired performance on a reaction time task plus reports of fatigue and discomfort after exposure for a 3-hour period to a 115-db. sound (Miles, 1953); decreased vigilance under a 100-db. noise condition (Broadbent, 1951, 1953); reports of irritation, distraction, and discomfort as a result of 30-minute periods of 111-db. noise (Miller, 1951).

Some physiological effects of noise are a rise in blood pressure following a loud unexpected sound (Lovell, 1941), a decrease in peristaltic contraction and flow of gastric juices following 10-minute periods of noise (Smith & Laird, 1930), and an increase in pulse rate during periods of noise while taking an intelligence test (Corso, 1952). Increased palmar sweating, EMG, and respiration amplitude were reported along with decreased pulse, finger volume, and respiration rate, and initial increase followed by a decrease in pulse rate (Davis, Buchwald, & Frankman, 1955). Lovett Doust and Schneider (1952) showed that oxygen saturation levels of the blood varied consistently with rhythmic auditory stimuli in both hospital patients and normals. Some effects of decreased oxygen saturation or anoxemia reported were poor concentration, tiredness, irritation, and dizziness. In another paper (Lovett Doust, 1953) it was pointed out that these findings of rhythmic effects upon oxygen saturation level might be applicable to mental illness where one of the characteristic features is relative anoxemia. This problem was considered in another paper by Lovett Doust and Schneider (1954) where some data was presented on "Rhythmic Sensory Bombardment Therapy." They formulated a hypothesis that relative anoxemia is one of the critical physiological spheres of mental illness, either in a resting state or in

stress situations. This approach stresses physiological effect of repetitive stimuli, in relation to different sense modalities, upon treatment of mental illness. The degree of actual effectiveness of rhythmic stimulation as yet has not been determined.

If the WS acts as an analgesic (Cook, 1960), then it is plausible that the UCS should not produce as strong a reaction under the WS condition when compared to the condition where WS is absent (N-WS), since the GSR varies with strength of effectiveness of the UCS (Furer & Hardy, 1950; Hovland & Riesen, 1940). Also, conventional anesthetics tend to raise the resistance level of the skin (Lindsley, 1951). However, if the WS does not act as an analgesic but as any other noise, then one would expect a greater GSR from the WS group. This proposition is based on the evidences that when the noise is noxious it tends to lower basic resistance level and it also functions to increase the response to shock by generating an increase in the activation level of the subject (Lacey, 1959). Respiration, like the GSR, has been used as a measure of reaction to a stressful situation (Ax, 1953; Schachter, 1957).

More specifically, the purpose of the present study was to determine the effect of WS upon the galvanic skin response to electric shock and upon frequency of respiration in normals and schizophrenics.

## METHOD

*Subjects.* Subjects for the experiment were 40 chronic undifferentiated schizophrenics who were patients at the Veterans Administration Hospital, Tomah, Wisconsin. The schizophrenic group had a mean age of 36.52 and a mean length of hospitalization of 61.35 months. All patients receiving any type of psychotropic drugs were taken off their medication 30 days prior to the experimental procedure. Also, 40 employees at the same hospital were used as normal controls. Both groups excluded the subjects under 25 or over 50 years of age. The mean age of normals was 35.78 years. These subjects were divided into four groups: schizophrenic-WS (Schiz-WS), schizophrenic-no WS (Schiz-NoWS), normal-WS (Nor-WS), and normal-no WS (Nor-NoWS).

*Apparatus.* The GSR and respiration were recorded with a multipurpose polygraph recorder. This apparatus recorded resistance to the nearest 1,000 ohms. Two GSR zinc electrodes used were of the finger clamp type and were placed on the index

and ring fingers of the right hand. The fingers were cleaned with alcohol solution and EEG electrode jelly was applied prior to attachment of electrodes. The electrodes provided contact with both of the finger pads. The GSR current applied to electrodes was a constant 20 microamperes, dc, which was then amplified and recorded on a pen polygraph. The current was regulated by an amplifier unit which had a self-contained regulated power supply for dc voltages. The subject resistance control on a preamplifier was used to balance basal resistance and had a dial for reading the electrode basal resistance. The subject was instructed to get his right hand and arm in a comfortable position on the arm of the chair and then let it go limp and refrain from moving it.

The UCS was an electric shock delivered by an Ensco Model PC-1 Portable Conditioner through zinc strap electrodes attached (electrode jelly) to the subject's left wrist. The shock was regulated by dc voltage stimulus which could be varied from 0 to 150 volts and was current regulated. The time duration of pain stimulus was set manually by the use of variable cams. The CS was a 15-watt light bulb behind an 8 × 12-inch milk glass screen placed 7 feet in front of the subject at eye level. The CS was presented for 5 seconds and was followed immediately by the UCS for 2 seconds. The duration of the CS and the onset of the UCS were controlled automatically by use of a cascaded multiple-phase electronic timer. The duration of the UCS was controlled in the same manner.

The WS and music that was used for attention was produced on the same record. The sound was reproduced with a stereophonic system consisting of a 40-watt amplifier and 2 sets of speakers. One of the speaker systems was placed 4 feet directly behind subject's chair and the other speaker was placed 3 feet in front of the subject. By means of the balance control on the amplifier, the sound was channeled as much as possible to the speaker behind the subject. The intensity of the sound and music was at continuous 70 db. This was not a subjective level, but a standard intensity level at the point where subjects were seated. The experimental room was air conditioned and during the actual running of the subjects the room was darkened except for an experimenter's lamp of low illumination (15 W).

**Procedure.** The subject was seated in a comfortable easy chair and the electrodes were attached. While the electrodes were connected, the subject was told in a conversational manner something of the nature of the experiment and what was expected of him. The following points were always covered:

1. The purpose of the experiment was to determine the effect of small electric shocks on his wrist upon sweating.

2. The only thing for the subject to do was to relax as much as possible and to refrain from moving about in the chair.

3. The WS groups were also told that to help them relax, some specially designed noise and

music would be played. Also, to be sure they got the full benefit of the noise and music they were told that they must listen to it by trying to follow the music through the noise.

The subjects were then given a series of shocks of ascending intensity beginning at 0 and they were instructed to tell the experimenter at which point the shock was "uncomfortable but not painful." When this level was determined the lights were extinguished and for the WS groups the sound with music was turned on. There was no difference between the levels of shock applied to the two groups; the mean shock value was 4.62 and 4.79 milliamperes for schizophrenics and normals, respectively.

Five minutes were then allowed to elapse for the subject to become acclimated to the situation, after which time the first base level of the GSR was determined. As soon as the base level was set, the first adaptation trial began.

There were 30 trials in all: 10 adaptation trials with the light alone, and 20 conditioning trials with light and shock paired. In order to minimize anticipation, the intertrial interval was 40, 50, or 60 seconds, each interval being chosen randomly with the qualification that each interval was used either 9 or 10 times.

## RESULTS

The means and standard deviations of respiration and GSR data are shown in Table 1. Due to the nature of these data, a nonparametric Mann-Whitney *U* test was applied to test the significances of differences between the groups.

**GSR.** The GSR scores were measured as changes in conductance. The conditioned response was the difference between the level at the point where the light came and the highest point for the duration of the light. The UCR was the difference between the level at the onset of shock and the highest point during 5 seconds following the onset of shock. All independent between-group comparisons

TABLE 1  
RESPIRATION RATES AND MAGNITUDE OF CONDUCTANCE CHANGES FOR SCHIZOPHRENIC AND NORMAL SUBJECTS

Subjects	Respiration		UCR		CR	
	M	SD	M	SD	M	SD
Nor-WS	10.28	2.23	39.16	18.62	27.38	10.03
Nor-NoWS	10.41	3.18	24.43	7.85	13.91	3.97
Schiz-WS	14.03	6.07	21.94	10.03	8.72	6.26
Schiz-NoWS	10.69	4.09	21.56	12.68	7.04	5.84

TABLE 2  
GSR AND RESPIRATION GROUP COMPARISONS

Conditions	Respiration		GSR			Base Levels	
	Adaptation	CR	Adaptation	CR	UCR	Initial	Terminal
Normals—white sound versus no white sound							
Overall	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
Last 10 trials				<i>ns</i>	$U = 22$ $p < .05$		
Last 5 trials			<i>ns</i>	$U = 23$ $p < .05$			
Schizophrenics—white sound versus no white sound							
Overall	$U = 113.5$ $p < .02$	$U = 101$ $p < .02$	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
Normals versus schizophrenics—white sound							
Overall	$U = 45$ $p < .02$	$U = 47$ $p < .02$	$U = 37$ $p < .02$	$U = 50$ $p < .05$	$U = 45$ $p < .02$	<i>ns</i>	<i>ns</i>
Normals versus schizophrenics—no white sound							
Overall	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>

were made using the Mann-Whitney  $U$  test. All within-group comparisons were made with the Wilcoxon Signed Ranks test.

**Adaptation.** The Nor-WS group was consistently above the other three groups in GSR. As can be seen from Table 2, this difference was significant only between the Nor-WS and Schiz-WS. The Nor-WS, however, had higher variance as compared with the Nor-NoWS group ( $F = 12.47$ ,  $p < .01$ ).

**Conditioned response.** The CRs for the four groups are shown in Figure 1. Using the last 5 versus the first 5 conditioning trials, the difference for the normal groups was significant ( $p < .01$ ). Also, as can be seen from Figure 1, the normal groups gave much greater CRs than did the schizophrenic patients who did not condition. These differences were significant at the .05 level of

confidence. In addition, the CRs of the Nor-WS group were maintained at a higher level than those of Nor-NoWS throughout conditioning, even though both groups tended to diminish in CR. This difference was found to be significant for the last 5 trials ( $p < .05$ ).

**Unconditioned responses.** Magnitude of UCRs is shown in Figure 2 for the four groups. Here again, the Nor-WS group is consistently above the other three groups. When this group is compared with the Nor-NoWS group, the difference is significant at the .05 level for the last 10 trials. The difference over all trials between the Nor-WS group and the Schiz-WS group was also significant at the .02 level.

**Base levels.** Two base levels were determined for each subject. One just preceded the first adaptation trial and the second followed



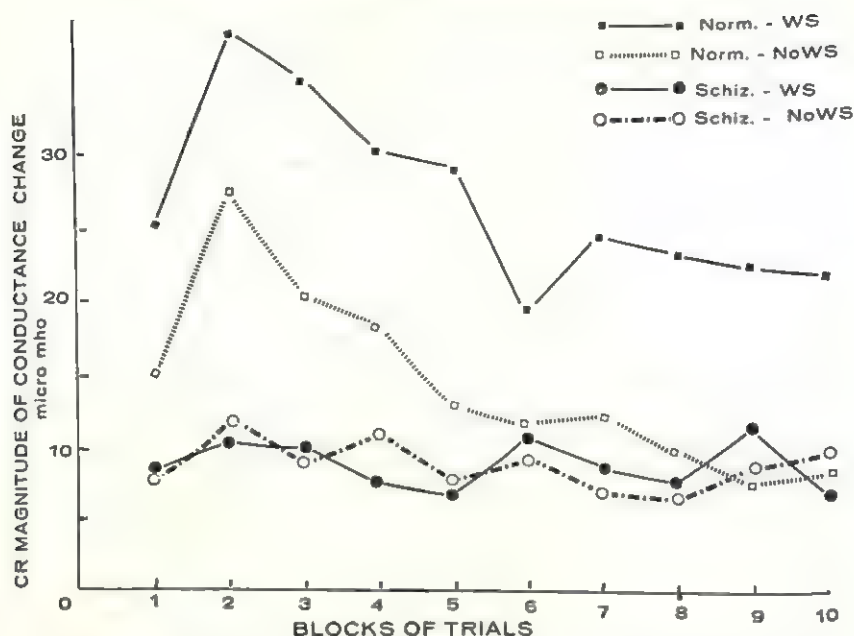


FIG. 1. Changes in conductance in CR for blocks of 2 trials each following 10 adaptation trials. (Each point is based on the mean value for 20 subjects.)

the last conditioning trial by 30 seconds. The comparisons that were made can be seen in Table 2, and none were significant. First versus final base level GSR differences were tested and the only significant difference found was for the Nor-WS group with the final base level conductance being higher than the initial level ( $p < .05$ ).

*Respiration.* The frequency of respiration was determined by the number of peaks on the respiration record for 20 seconds preceding the CS and 20 seconds following the onset of the CS for each trial. The results of this measure are shown in Figure 3. Here, the Schiz-WS group is consistently higher than the other three groups. The difference be-

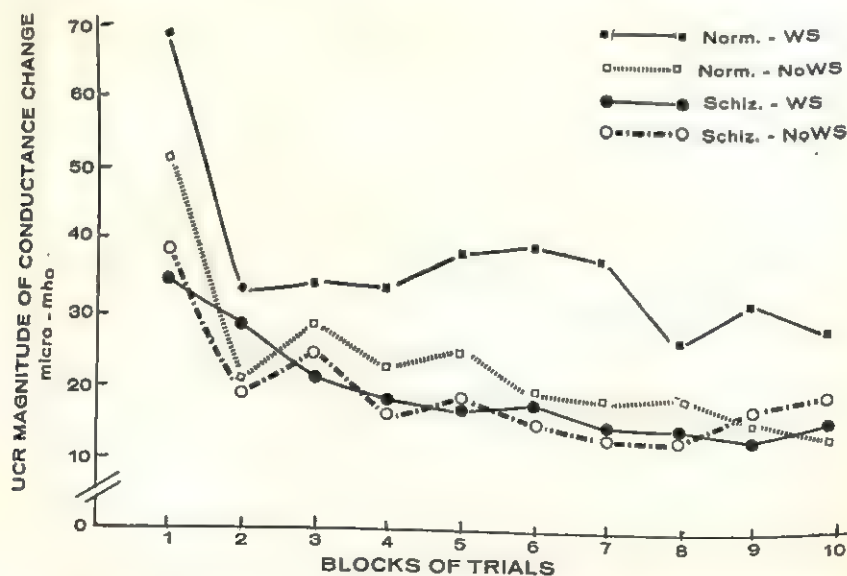


FIG. 2. Changes in conductance in UCR for blocks of 2 trials each following 10 adaptation trials. (Each is based on the mean value for 20 subjects.)

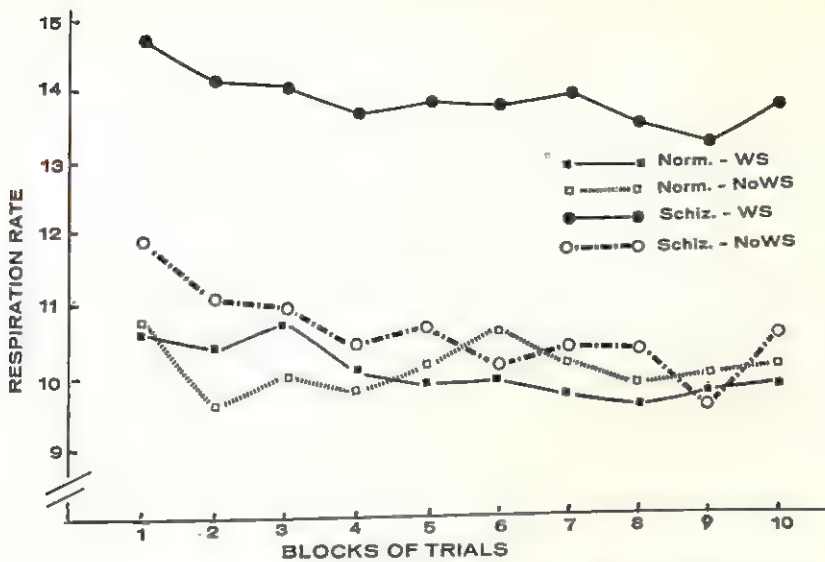


FIG. 3. Respiration rate for blocks of 2 trials each following 10 adaptation trials. (Each point represents mean frequency of respiration of 20 subjects for 20 seconds preceding and following CS.)

tween the Schiz-WS and Schiz-NoWS groups was significant ( $p < .02$ ) for adaptation and conditioning trials. Furthermore, when Schiz-WS and Nor-WS groups were compared, significant differences were revealed for both the adaptation and the conditioning periods ( $p < .02$ ).

#### DISCUSSION

In comparing the effects of WS in this study with the effects observed in the dentist's chair, several points of difference in the two situations must be recognized. First, only one intensity of WS was used in this experiment, while the dental patient selects his own intensity level and can change it at any time. Moreover, the WS in this experiment was continuous. In the applied situation, the patient can manipulate the sound and let it run continuously or in short bursts, if so desired. Another difference is that of the instructions. The dental patients were most certainly told that the WS would mask their pain while the subjects of this experiment were merely told that the WS would help them to relax. Also, the mere fact that the dental patient controls the sound not only gives him some control over the situation, but this little task may tend to occupy the patient and serve as a distraction from the

dentist's operations. On the other hand, in the present experimental setting the subject had only to sit in a chair for approximately 40 minutes without any activity.

For the GSR data, the most significant result is that when there was a difference between WS and NoWS groups, the WS group always produced a greater response. This is the opposite of what would be predicted if WS were to function as an analgesic. This finding seems to warrant the conclusion that, in this experiment, WS was a noxious stimulus much like any other continuous loud noise. This would indicate that any possible analgesic effects of WS might be due not to any somatic consequences of WS per se, but to the effects of another noxious stimulus which may serve to distract the patient from other noxious or painful stimuli. Equally important was the failure of schizophrenics to show GSR conditioning, whereas there was rapid conditioning in the normal group which was followed by a swift adaptation or extinction (Figure 1). This extinction may have been due to the fact that the UCS was energized immediately following termination of CS. Thus, this timing factor may have produced some fusion between UCR and CR. Although there are no adequate means to explain this specific difference between the

two populations, it is noteworthy that schizophrenics' GSR responsivity was generally suppressed as compared to normals.

Respiration data reveal some unexpected but interesting trends. Here again, the WS group gave a stronger response, but with this measure it was the schizophrenic group that showed significant increase and not the normal group, as was the case with the GSR data. An interesting hypothesis, following the ideas of Lovett Doust and Schneider (1952), that noise causes the oxygen saturation level of the blood to decrease and thus triggers the response of an increased respiration rate, would be that oxygen saturation was lowered in schizophrenics under WS condition. Respiration rate is, of course, only an indirect measure of oxygen level which was directly measurable by Lovett Doust and Schneider.

The major finding of this study was the differential response of schizophrenics and normals to WS, as measured by GSR and respiration rate. Although there are some inconsistencies in other related data, there is some evidence of response patterns indicating distinguishable connection to experiences of either fear or anger. Ax (1953) found a significantly higher number of GSRs in anger than fear conditions. Although the present study utilized magnitude rather than number of GSRs, it was assumed that for the present short-term durations, magnitude would be a more appropriate measure in reflecting levels of activation. Hence fear-anger differentiation, as demonstrated by Ax, is inferred in explanation of the present findings. Secondly, respiration rate was found to increase significantly more in fear than in anger by both Ax (1953) and Schachter (1957). Unfortunately, the present investigators did not elicit any subjective reports regarding the subjects' experiences during the experiment. It was not anticipated in the original design that group differences would emerge with respect to respiration and GSR. Consequently, in this study, there is no clear-cut evidence of either fear or anger as a predominant experience.

Generally, GSR conditioning rate is associated with increase in severity of anxiety (Schiff, Dougan, & Welch, 1949). In ad-

dition, there is evidence that respiration rate is higher for schizophrenics than for normals under stress-like conditions (Jurko, Jost, & Hill, 1952). Kryter (1950) has pointed out that different kinds of sounds may have influence on feelings. He stressed that irregular sounds tend to be more annoying than steady cues. Since WS waves are of varying intensity, resembling waves in water, it is not surprising that a form of annoyance would result from it. Regardless of fear-anger and/or oxygen saturation speculations, it was evident that *the physiological responses to white sound produced in this experiment were clearly distinguishable between schizophrenics and normals.*

Since respiration is under greater voluntary control than GSR, which is a purer manifestation of autonomic response, it is plausible that normals exercised greater voluntary control to the WS stimulation although their response may have been characteristic of anger or annoyance while schizophrenics' responses were channeled through fear. This would be characteristic of generalized lack of schizophrenics' emotional control.

In addition, the possibility that some degree of relative anoxemia is involved in mental illness may serve as an additional explanation for increased respiration rate by schizophrenics. Lovett Doust and Schneider (1954) postulated that decrease of oxygen saturation is likely to be one of the critical characteristics of mental illness. In view of the present findings, varying levels of WS intensity and different types of instructions suggesting anticipated effects of WS should be manipulated in clarifying presumed audio analgesic property of WS. Relationship of GSR and respiration rate should be tested under conditions of WS within situations which could be clearly distinguished as either fear or anger producing in order to follow up present propositions. Although the instructions and the sensory input were identical for both groups, either differential perception or discriminable responsivity between schizophrenics and normals was demonstrated. At this point the most plausible explanation for present results is conceived in terms of fear-anger constructs and oxygen saturation levels. Within the foregoing analysis it must



be pointed out that the fear-anger distinction is a tentative one, requiring caution against general acceptance without more thorough investigation of subtleties in autonomic responsivity associated with different states of arousal of schizophrenics and normals. Oxygen saturation must also be explored in relation to physiological correlates of white sound.

## REFERENCES

- AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, 15, 433-442.
- BROADBENT, D. E. Noise, paced performance and vigilance tasks. *Med. Res. Council, Appl. Psychol. Res. Unit Rep.*, 1951, No. 165/51.
- BROADBENT, D. E. Noise, paced performance and vigilance tasks. *Brit. J. Psychol.*, 1953, 44, 295-303.
- COOK, E. *Audio analgesia*. Stamford, Conn.: Cook Laboratories, 1960.
- CORSO, J. F. The effects of noise on human behavior. *USAF WADC tech. Rep.*, 1952, No. 53-81.
- DAVIS, R. C., BUCHWALD, A. M., & FRANKMAN, R. W. Autonomic and muscular responses and their relation to simple stimuli. *Psychol. Monogr.*, 1955, 69(20, Whole No. 405).
- DUNCKER, K. Some preliminary experiments on the mutual influence of pains. *Psychol. Forsch.*, 1937, 21, 311-326.
- FURER, M., & HARDY, J. D. The reaction to pain as determined by the galvanic skin response. *Res. Publ. Ass. nerv. ment. Dis.*, 1950, 29, 72-89.
- GARDNER, W. J., & LICKLIDER, J. C. R. Auditory analgesia in dental operations. *J. Amer. Dent. Ass.*, 1959, 59, 1144-1149.
- HOVLAND, C. I., & RIESEN, A. H. Magnitude of galvanic and vasomotor response as a function of stimulus intensity. *J. gen. Psychol.*, 1940, 23, 103-121.
- JURKO, M., JOST, H., & HILL, T. S. Pathology of the energy system: An experimental-clinical study of physiological adaptive capacities in a non-patient, a psychoneurotic, and an early paranoid schizophrenic group. *J. Psychol.*, 1952, 33, 183-198.
- KRYTER, K. D. The effects of noise on man. *J. speech hear. Dis. Monogr. Suppl.*, 1950, No. 1.
- LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959. P. 161.
- LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. London: Chapman & Hall, 1951. P. 475.
- LOVELL, G. D. Physiological and motor responses to a regularly recurring sound. *Psychol. Bull.*, 1941, 38, 715. (Abstract)
- LOVETT DOUST, J. W. Physiological factors underlying therapeutic success in mental disorders. *Canad. Med. Assoc. J.*, 1953, 69, 108-114.
- LOVETT DOUST, J. W., & SCHNEIDER, R. A. Studies on the physiology of awareness: The effect of rhythmic sensory bombardment on emotions, blood oxygen-saturation and the levels of consciousness. *J. ment. Sci.*, 1952, 84, 640-653.
- LOVETT DOUST, J. W., & SCHNEIDER, R. A. Rhythmic sensory bombardment therapy (R.S.B.T.). *Dis. nerv. Sys.*, 1954, 15, 3-15.
- MALMO, R. B., SHAGASS, C., & DAVIS, J. F. Electromyographic studies of muscular tension in psychiatric patients under stress. *J. clin. exp. Psychopath.*, 1951, 12, 45-66.
- MILES, W. R. Immediate psychological effects. In, *Benox report: An exploratory study of the biological effect of noise*. Contract Report No. N6, Dec. 1, 1953, University of Chicago, Contract ORI 020. Pp. 81-95.
- MILLER, H. G. A study of the interference effects of high intensity sound on the retention of visually and auditorily learned material. Unpublished doctoral dissertation, Pennsylvania State College, 1951.
- MILLER, J. G. *The development of experimental stress-sensitive tests for predicting performance in military tasks*. Washington, D. C.: Psychological Research Association, 1953.
- SCHACHTER, J. Pain, fear and anger in hypertensives and normotensives. *Psychosom. Med.*, 1957, 19, 17-29.
- SCHIFF, E., DOUGAN, C., & WELCH, L. The conditioned PGR and the EEG as indicators of anxiety. *J. abnorm. soc. Psychol.*, 1949, 44, 549-552.
- SMITH, E. L., & LAIRD, D. A. The loudness of auditory stimuli which affect stomach contractions in healthy human beings. *J. Acoust. Soc. Amer.*, 1930, 2, 94-98.

(Received November 13, 1962)

## AFFECTIVE STIMULI AND DISTURBANCE OF THOUGHT PROCESSES<sup>1</sup>

GAYLE DEERING<sup>2</sup>

*Teachers College, Columbia University*

The study investigated the effects of affectively toned and neutral adjectives on the incidental memory functioning of matched groups of 40 schizophrenic and normal female Ss. The adjectives which had been matched for familiarity and rated by judges were presented in a word association test in which the number of different associations, association recall, and stimulus word recognition scores were measured. Conclusions are: (a) schizophrenic Ss gave a larger number of different associations to affective words than normal Ss (.01 level); (b) there was no evidence that schizophrenic Ss, in comparison with normal Ss, recalled relatively more associations to neutral words than to affective words; (c) there was no evidence that schizophrenic Ss recognized stimulus words as well as normals.

Schizophrenia has been characterized as a disturbance in thought processes which often results in noncommunicative language. Theoretical formulations and clinical observations suggest that affective stimuli impair schizophrenic patients' thinking (Arieti, 1955; Bleuler, 1950; Fromm-Reichmann, 1959). Rapaport, Gill, and Schafer (1946) have elaborated upon the theory that the thinking of schizophrenic patients is easily disturbed. According to them, "good adjustment is characterized by personal and/or traumatic material not being mobilized easily, either in life or in the testing situation; while in maladjustment, such material is often mobilized even in situations which are only remotely relevant [p. 18]." Their theoretical formula-

tion implies that schizophrenic patients show a greater disturbance in thought processes in dealing with affective material than with neutral material.

Considerable research has accumulated supporting the hypothesis that affective stimuli are particularly disturbing to schizophrenic patients. Much of the research investigated complex conceptual thinking (Tomblen, 1957; Wexler, 1955; Whiteman, 1954). The main criticism of these studies has been the lack of equating the affective stimulus material with the neutral stimulus material for difficulty. In previous studies, the poorer performance of schizophrenic patients in responding to affective stimulus material may have been due to its greater difficulty level rather than to its affective content. The continuing importance of the hypothesis, together with the conflicting interpretations of past research, suggests further exploration with a research design that avoids complex cognitive tasks.

Incidental memory appears to be a relatively simple cognitive function that does not place schizophrenic patients at a disadvantage and was selected for exploration. So far, few research studies have investigated the disruptive influence of affective stimuli on schizophrenic incidental memory directly. The present study explored the effects of affective stimuli on schizophrenic association, recall, and recognition responses to an experimental word association test.

<sup>1</sup> The article is based upon a dissertation submitted to Teachers College, Columbia University, in partial fulfillment of the requirements for the degree of doctor of philosophy. Grateful acknowledgment is extended to members of the dissertation committee, Laurance F. Shaffer (Chairman), Joel R. Davitz, and William N. Thetford, and to Rosedith Sitgreaves. In the collection of data, the writer is indebted to the staff members and patients at the following hospitals: City Hospital, Elmhurst, New York; Hillside Hospital, Glen Oaks, New York; Jacobi Hospital, Bronx, New York; Massachusetts Mental Health Center, Boston, Massachusetts; New York State Psychiatric Institute, New York, New York; St. Vincent's Hospital, New York, New York.

<sup>2</sup> Research associate, Institute of International Education, Foreign Physicians' Training Study, Goldwater Memorial Hospital, Welfare Island, New York, New York.



### *Disturbance Indicators in Word Association and Related Research*

The word association test was selected as the research instrument because it elicits information about memory functioning while avoiding complex memory tasks. Evidence of association and recall disturbance was based on the following formulation of disturbance on word association tests. In associating to a stimulus word, an individual without emotional maladjustment generally responds to a word association test with conventional associations conceptually related to the stimulus word (Rapaport, Gill, & Schafer, 1946). In the emotionally disturbed individual, deviations from the conventional-conceptual mode of responding are to be expected and are indicators of association disturbance.

Good recall indicates stable thought organization which, according to Rapaport, Gill, and Schafer (1946), is characteristic of normal individuals; whereas poor recall suggests instability of thought organization which is characteristic of severely disturbed individuals and indicates a disturbance in recall.

In a study which investigated association and recall disturbances in schizophrenic, neurotic, and normal individuals, Rapaport, Gill, and Schafer found evidence of disturbance in the schizophrenic, neurotic, and normal individuals in responding to affective stimulus words. Their findings are inconclusive because the stimulus words had not been equated for familiarity and the patient groups had not been matched with the control group. Jones (1957) who exercised greater care in matching his groups and controlled for word familiarity found schizophrenic patients gave more unique associations to affective stimulus words than normals.

Recognition is another aspect of memory. Because recognition of stimulus words does not require the subject to recall his own responses, it was assumed in the study that recognition of stimulus words, in contradistinction to recall of associations, measured a lesser degree of personal involvement and active participation. Recognition was considered an easier task than recall of associations. No reported word association studies have investigated the recognition of stimulus

words by schizophrenic subjects. However, several tachistoscopic studies have investigated the recognition of stimulus words. McGinnies and Adornetto's (1952) and Ressler's (1950) findings suggest that schizophrenic and normal subjects do not recognize affective stimulus words as well as neutral words. However, the authors did not match the affective words with the neutral words for familiarity. In Newberry's study (1954), in which stimulus words equated for familiarity were presented tachistoscopically, neither speed of recognition nor recall of affectively toned words differentiated between the schizophrenic and normal subjects. Failure to find a difference in recognition of stimulus words between schizophrenic and normal subjects suggests that schizophrenic subjects perform as well as normal subjects in a recognition task in which personal, active involvement is at a minimum.

Theoretical formulations and research findings suggest that schizophrenic subjects give more different associations and recall fewer associations to affective stimulus words than to neutral words than do normal subjects; schizophrenic subjects recognize affective and neutral stimulus words as well as nonschizophrenic subjects when the stimulus words are controlled for familiarity.

### *Hypotheses*

1. In word association, schizophrenic patients, in comparison with nonpatients, give a greater number of different associations to affective stimulus words.
2. In recall of associations, schizophrenic patients, in comparison with nonpatients, recall relatively more associations to neutral words than to affective words.
3. In recognition of stimulus words, schizophrenic and nonpatient individuals show no difference in recognition of affective and neutral stimulus words.

Although no specific hypotheses were stated regarding educational level, the effects of education were investigated.

### METHOD

#### *Subjects*

Forty first-admission schizophrenic patients hospitalized less than 11 months with a medical diag-



nosis of schizophrenia formed the patient group. None of the schizophrenic subjects had received shock within the previous 6 months or had been diagnosed as having organic impairment.

Forty nonpatient subjects who appeared able to function adequately in their vocational role and who were not receiving psychotherapy comprised the normal group. All subjects were United States born white females between the ages of 21 and 35 years; one half were high school graduates, the remaining one half had some college education. Nonpatient subjects' vocations were roughly matched with schizophrenic subjects' vocations prior to hospitalization. The mean ages for the high school groups were 28.05 for schizophrenic and 27.25 for normal; for the college groups, 25.65 for schizophrenic and 25.20 for normal. The vocabulary raw scores on the Shipley Institute of Living Test for the high school groups were 31.70 for schizophrenic and 32.05 for normal; for college groups, 34.45 for schizophrenic and 35.50 for normal. The differences between the diagnostic categories for the high school and college groups separately on these variables appear to be nonsignificant.

### *Procedure*

Each subject was seen individually by the same experimenter. The experimental word association test was administered according to standard procedure. Upon completion of the test, each subject was asked to recall her original associations as the experimenter repeated the stimulus words which had been randomized a second time. Following recall, the recognition test of the stimulus words was administered. Each subject was asked to encircle those words which had been read to her by the experimenter. The vocabulary part of the Shipley Institute of Living Scale was administered last.

### *Experimental Word List*

Adjectives equated for frequency from 1 to 25 times per million words (Thorndike & Lorge, 1944) had been rated by 20 inpatient and 20 normal judges to determine the affective and neutral values of the words. The judges, who were matched on age, sex, and education with the patient and normal groups, rated 175 adjectives on a 7-point scale ranging from very pleasant to very unpleasant. The rating scale was administered individually twice over a 1-week interval. Eighty adjectives<sup>3</sup> judged initially

<sup>3</sup> GIFTED, GRACEFUL, LOVING, SKILLFUL, YOUTHFUL, ALERT, JOYFUL, RESTFUL, SINCERE, LOYAL, TACTFUL, THANKFUL, WITTY, FRUITFUL, HOPEFUL, TRUSTING, TRUTHFUL, GALLANT, HELPFUL, CORDIAL, BRUTAL, ROTTEN, UNJUST, HARMFUL, LONESOME, PAINFUL, SELFISH, STUPID, THOUGHTLESS, DREARY, FILTHY, HATEFUL, JEALOUS, WORTHLESS, BARREN, BLOODY, CALLOUS, FRIENDLESS, SCORNFUL, VULGAR, CONCAVE, EARTHEN, FRECKLED, LANKY, METRIC, RANDOM, NEUTRAL, OPTIC, OVAL, POROUS, CUBIC, BUSHY, COUNTLESS, FISCAL, FOCAL, FLEETING, MASSIVE, NIGHTLY, POLAR, SPIRAL, SPOTTED, GRAPHIC, HUSKY, LENGTHY, PLURAL,

as pleasant, unpleasant, or neutral by at least 65% of all judges comprised the experimental word list. An overall reliability coefficient of .89 based on the subjects' test-retest ratings suggests that the ratings had a high degree of stability. The final list contained five dummy adjectives at the beginning and end of the list.

### *Experimental Recognition Test*

For the recognition phase of the experiment, 20 affective and 20 neutral stimulus words were randomly selected and embedded in a list of 120 new adjectives. A measure of test reliability was calculated by the split-half method. With the Spearman-Brown correction, a reliability coefficient of .82 was obtained on the embedded adjectives.

## RESULTS AND CONCLUSIONS

The results are presented in terms of idiosyncratic association, association recall, and stimulus word recognition scores. All scores were group scores measured across words. The variances of these scores were found to be homogeneous. Certain trends in the data which appeared significant, even though they had not been hypothesized, are reported.

### *Idiosyncratic Associations*

The mean idiosyncratic association scores refer to the mean number of different associations given to the 40 affective and 40 neutral words by the subjects in the schizophrenic and normal groups. High idiosyncratic association scores for a group indicate an abundance of different associations. Table 1 gives the means and standard deviations of the idiosyncratic association scores for affective and neutral stimulus words.

In terms of quantitative scores, the first hypothesis predicted that affective stimulus words elicit a greater number of different associations from subjects in schizophrenic groups than from subjects in normal groups. An analysis of variance revealed an *F* value of 24.47, significant beyond the .01 level, between diagnostic categories. Hypothesis 1 was supported. There was no evidence that educational level influenced the number of different associations for either the schizophrenic or normal subjects.

In analyzing the data on associations to

SPINAL, SPONGY, SLANTING, WOOLY, BULKY, CIVIC, COSMIC, ELDEST, FLAKY, FOAMY, GRAINED, NATAL, OLDEN, TRIBAL, YEARLY.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF IDIOSYNCRATIC ASSOCIATION SCORES FOR AFFECTIVE AND NEUTRAL STIMULUS WORDS

Group and education	Affective words		Neutral words	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Schizophrenic				
High school	13.42	3.10	10.98	3.54
Normal				
High school	10.78	3.30	10.42	3.30
Schizophrenic				
College	13.82	2.88	10.90	3.20
Normal				
College	11.85	2.48	10.32	2.92

Note.—There are 40 words in each category and 20 subjects in each group.

neutral words by analysis of variance, there were no differences found for idiosyncratic association scores for neutral words between schizophrenic and normal groups nor between educational levels.

### Association Recall

The mean association recall scores refer to the mean number of correctly recalled associations given to the 40 affective and 40 neutral words by the subjects in the schizophrenic and normal groups. Table 2 gives the means and standard deviations of association recall scores.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF ASSOCIATION RECALL SCORES FOR AFFECTIVE AND NEUTRAL STIMULUS WORDS

Group and education	Affective		Neutral		Combined	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Schizophrenic						
High school	10.90	2.96	15.58	2.75		
College	12.15	2.96	16.45	2.21		
Combined					27.54	6.56
Normal						
High school	13.10	3.49	16.78	2.58		
College	13.08	3.50	17.47	2.21		
Combined					30.21	6.85

Note.—There are 40 words in each category and 20 subjects in each group.

The second hypothesis predicted that the subjects in the schizophrenic groups recall relatively more associations to neutral words than to affective words in comparison with the subjects in normal groups. To test the second hypothesis, it was necessary to investigate the interaction between type of word and diagnostic category. The results of an analysis of variance found this interaction to be nonsignificant. Hypothesis 2 was not upheld.

Differences at the .01 level of significance were found between types of words, diagnostic categories, and educational levels. These differences show that more associations to neutral words were recalled than associa-

TABLE 3

MEANS AND STANDARD DEVIATIONS OF RECOGNITION SCORES FOR AFFECTIVE AND NEUTRAL STIMULUS WORDS

Group and education	Affective		Neutral		Combined	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Schizophrenic						
High school	13.60	2.58	15.90	2.84		
College	15.20	2.02	16.70	2.79		
Combined					30.70	4.95
Normal						
High school	14.95	2.95	17.55	1.76		
College	16.10	3.40	17.95	2.01		
Combined					33.28	5.12

Note.—There are 20 words in each category and 20 subjects in each group.

tions to affective words by subjects in all groups; more associations were recalled by subjects in normal groups than by subjects in schizophrenic groups; and more associations were recalled by subjects in college groups than by subjects in high school groups.

### Stimulus Word Recognition

The mean recognition scores refer to the mean number of times the 20 affective and 20 neutral stimulus words were recognized by subjects within the schizophrenic and normal groups. The means and standard deviations are given in Table 3.

The third hypothesis predicted that schizophrenic individuals recognize affective and neutral stimulus words as well as normal individuals. To test Hypothesis 3, it was



necessary to investigate the effect of diagnostic categories on recognition scores. An analysis of variance resulted in an  $F$  value of 22.25, significant beyond the .01 level, for diagnostic categories. Subjects in schizophrenic groups did not recognize affective and neutral stimulus words as well as subjects in normal groups. Hypothesis 3 was not upheld.

Differences at the .01 level of significance were found between types of words and educational levels. The findings show that neutral stimulus words were recognized more frequently than affective stimulus words by the subjects in all groups; and more stimulus words were recognized by the subjects in the college groups than by the subjects in the high school groups.

### DISCUSSION

The present finding showing that schizophrenic subjects, but not normal subjects, increased their idiosyncratic associations to affective stimulus words is in agreement with Jones' (1957) findings. This increase in schizophrenic idiosyncratic associations to affective stimulus words appears to define affectively toned material as a variable which influences schizophrenic associations.

Similar to the results reported by Rapaport, Gill, and Schafer (1946), the present study demonstrated that schizophrenic and normal subjects had poorer recall of associations to affective stimulus words than to neutral stimulus words. These findings suggest that Rapaport, Gill, and Schafer's theoretical formulation upon which the first two hypotheses were based should be re-evaluated. Rapaport, Gill, and Schafer based their formulation regarding the disruptive influence of sexual, anal, oral, familial, and aggressive stimulus words on the following assumption.

It appears then that stimulus-words related to those emotional areas which according to clinical psychology are the likely areas of conflict, do call into action the internally existing conflicts, as evidenced by the disruption of the association and reproduction processes [p. 83].

Can it be assumed that the affective stimulus words in the present study touched on areas of conflict because these stimulus words elicited more idiosyncratic associations from

the subjects in the schizophrenic groups than from the subjects in the normal groups and resulted in poorer recall for the subjects in all groups? This assumption appears to be based on circular reasoning which mitigates the usefulness of Rapaport, Gill, and Schafer's (1946) formulation. An alternative explanation is the possibility that affective stimulus words, even though they are equated for familiarity with neutral words, are capable of eliciting a greater number of associations than neutral words. This interpretation is further strengthened by the present finding of poorer recall of associations to affective words by normal subjects as well as schizophrenic subjects.

To support Rapaport, Gill, and Schafer's (1946) assumption, it would appear necessary to demonstrate that normal individuals, presumably free from emotional maladjustment and internally existing conflicts, recall their associations to affective stimulus words as well as those to neutral stimulus words; whereas schizophrenic individuals, who exemplify severe maladjustment, show more disruption in recall of associations to affective stimulus words than to neutral stimulus words. That normal and schizophrenic subjects demonstrated poorer recall of associations to affective stimulus words than to neutral stimulus words seems to indicate that the nature of the stimulus words rather than the nature of the subjects is responsible for the difference in recall scores.

Whether or not recall of associations to the affective stimulus words is related to the disruptive effects of affective content or to a plethora of possible associations to these stimulus words raises a question regarding experimental control. Laffal (1955) investigated response faults on words equated for familiarity. He found recall failures correlated .646 with the number of different associations given to the stimulus words. In the present study, recall failures correlated .334 with the number of different associations given to stimulus words. Although this correlation is considerably lower than Laffal's, it still demonstrates a positive relationship between recall failures and idiosyncratic associations. Laffal suggests that in order to evaluate the effect of emotional factors of



words, stimulus words with few associations should be selected. His suggestion seems relevant to the present study. Thus far, the only available control on words measures the frequency of occurrence in printed reading matter. It would appear desirable to construct an association frequency measure for words equated for familiarity along the lines set up by Laffal.

The present results on recognition are in disagreement with the tachistoscopic study in which familiarity was controlled (Newberry, 1954). The tachistoscopic study showed no difference in recognition between the affective and neutral stimulus words. The disagreement between these findings may be due in part to the tasks which the subjects were asked to perform. In the present study, the subject was not asked to attend carefully to the stimulus words, whereas in the tachistoscopic studies, the subject was requested to concentrate on the words. This difference in degree of attention may explain in part the findings. The present findings do not support the assumption that recognition ability is a relatively easy task which does not differentiate between schizophrenic and normal individuals.

#### REFERENCES

- ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.
- BLEULER, E. *Dementia praecox: Or, the group of schizophrenias*. New York: International Universities Press, 1950.
- FROMM-REICHMANN, FRIEDA. *Psychoanalysis and psychotherapy*. Chicago: Univer. Chicago Press, 1959.
- JONES, R. A. *Some effects of emotion on the use of words by schizophrenics*. Washington, D. C.: Catholic Univer. America Press, 1957.
- LAFFAL, J. Response faults in word association as a function of response entropy. *J. abnorm. soc. Psychol.*, 1955, 50, 265-270.
- MCGINNIES, E., & ADORNETTO, J. Perceptual defense in normal and schizophrenic observers. *J. abnorm. soc. Psychol.*, 1952, 47, 833-837.
- NEWBERRY, J. J. Perceptual thresholds of paranoid schizophrenics as a function of affectively toned words having interpersonal implications. Unpublished doctoral dissertation, Northwestern University, 1954.
- RAPAPORT, D., GILL, M., & SCHAFER, R. The word association test. In, *Diagnostic psychological testing*. Vol. 2. Chicago: Year Book Publishers, 1946. Pp. 13-84.
- RESSNER, J. E. The pre-recognition responses of schizophrenic patients and normals to the stimuli in a word association test. Unpublished doctoral dissertation, Columbia University, 1950.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.
- TOMBLEN, D. Anxiety and conceptual impairment in schizophrenia. Unpublished doctoral dissertation, Teachers College, Columbia University, 1957.
- WEXLER, F. The influence of emotional and non-emotional content upon conceptual thinking in schizophrenia. Unpublished doctoral dissertation, Teachers College, Columbia University, 1955.
- WHITEMAN, M. The performance of schizophrenics on social concepts. *J. abnorm. soc. Psychol.*, 1954, 49, 266-271.

(Received November 19, 1962)

## THE D 48 TEST AS A MEASURE OF GENERAL ABILITY AMONG GRADE SCHOOL CHILDREN

HARRISON G. GOUGH AND GEORGE DOMINO

University of California, Berkeley

The D 48 (or "Dominoes") test of intellectual ability is widely used in Europe, but is almost unknown in the United States. Because of its high loading on  $g$ , its ease of administration, and its potentiality as an instrument for use in cross-cultural comparisons, the test would appear to be an important candidate for study by American psychologists. In an initial study of 86 California school children it was found that D 48 item difficulties and mean scores were highly similar to data previously available in Europe. The test also showed satisfactory predictive validity (coefficients of  $+.58$  and  $+.45$ ) for scholastic achievement in the 5th and 6th grades. The results uphold the validity of the D 48, and warrant the hope that other researchers will undertake evaluations of the test.

Although widely used in Europe as a measure of intellectual ability, the D 48 or "Dominoes" test is more or less unknown in the United States. For example, two recent textbooks on testing (Anastasi, 1961; Cronbach, 1960) make no mention of the D 48, and the instrument is also uncited in Buros' *Tests in Print* (1961), *Fourth Mental Measurements Yearbook* (1953), and *Fifth Mental Measurements Yearbook* (1959). In Europe, however, the D 48 is a standard item in texts on measurement (cf. Callonghi, 1956; Pichot, 1949), and carries a lengthy research bibliography (e.g., Ferracuti & Rizzo, 1959; Pasquasy & Doutrepoint, 1956; Pichot, Rennes, & Taver, 1953).

Beyond the need of psychologists in the United States to be informed about current practices in Europe, there are additional reasons for paying attention to the D 48 test. One of these is the potentiality of the D 48 for cross-cultural measurement of ability. The items in the test consist of domino sequences, and hence draw on a background of experience found in nearly all literate societies. The game of dominoes is apparently as ancient in China as playing cards, for which specific references dating back at least 1,000 years may be found. The first record of the game in Europe was in Italy and France in the middle of the eighteenth century, and from there the game was transmitted to England. Anthropologists have also recorded

evidence of the game among the Inuit Eskimos of the Hudson Strait region.

In the game of dominoes, therefore, may be found kinds of stimuli and modes of conceptualizing which are universal or well-nigh universal. The ordinary approach in cross-cultural (or "culture free") testing is to search for stimuli which are new and/or unknown to all, so as to minimize differentials of past experience. An equally logical approach would be to utilize test stimuli which are known to everyone, and which draw on a considerable fund of childhood experience, so that familiarity with the stimuli and the conceptual modes employed in handling them may be equated. The D 48 test, being based on the game of dominoes, would seem to be a worthy candidate for cross-cultural testing under this second principle.

A second reason for attending to the D 48 test is that it is almost entirely nonverbal. The problems are nonverbal, and the instructions are so simple as to make almost no demand upon verbal ability. There are many testing situations in which nonverbal instruments are to be desired, and such instruments are by no means abundant in our catalogues of tests.

A third virtue of the D 48 test is that it is highly saturated with  $g$  (Vernon, 1950, p. 23). In fact, the loading of the D 48 on  $g$  (.87) was even higher than that of the Progressive Matrices (.79) in Vernon's anal-

ysis. Furthermore, the D 48 test was entirely free of the nuisance loading on the spatial factor which is a problem with the Matrices.

A fourth advantage of the D 48 test is that it is easy to administer and to score, and can be given in a brief period of time (30 minutes) either individually or to groups.

#### DEVELOPMENT OF THE TEST

The testing method was first developed in 1943 by Anstey and Illing,<sup>1</sup> for the Directorate for Selection of Personnel of the British War Office. The first test, called SP 11, contained 48 items. A revised 48-item version was prepared by Anstey, following World War II, for use by the National Institute of Industrial Psychology (NIIP) in London. Anstey also prepared two other versions, of 25 and 36 items each, for use by the British Civil Service Commission with university graduates. All four editions are in current use in England. A new experimental 42-item version in color has just been completed by Anstey.

A French adaptation of the NIIP edition of the Dominoes test was prepared in the late 1940's by Binois (cf. Pasquasy & Doutrepoint, 1956, p. 22), and published by the Centre de Psychologie Appliquée in Paris under the name "D 48." This version is also published in Italy by the Organizzazioni Speciali of Florence and was brought out in the United States in 1962 by the Consulting Psychologists Press (CPP) of Palo Alto. It is the CPP version which was used in the present study.

The D 48 test contains 44 problems, each consisting of a series of dominoes defining a principle of progression. The last domino in the series is left blank, and the subject is required to enter the proper numbers on each half of the domino. The principles involved in the progressions vary from problem to problem, and include simple addition, identities, identities with reversals, double progression, subtraction, subtraction with progression, etc. The sequence of items is in general from less to more difficult, with the

exception of five or six quite difficult items in the middle of the series.

Instructions and sample problems are given on the cover of the test booklet, and can be presented in 4 or 5 minutes. Testing time is set at 25 minutes, and most of the studies with the D 48 have observed this limit. However, the D 48 can be used as a power test and given without restriction on time (cf. Ferracuti & Rizzo, 1959).

#### PROCEDURE

The D 48 test was given to the 86 fifth and sixth grade students in a California elementary school.<sup>2</sup> Three specific goals were envisaged for the study: (a) to establish item difficulties and to relate these to similar data from studies in Europe, (b) to relate total scores to those from comparable European samples, and (c) to evaluate the validity of the D 48 as a predictor of scholastic achievement.

The percentage of students answering each item correctly is given in Table 1, along with the rank order for each item. The progression of difficulty is in general agreement with the item position, but there are exceptions. For example, Item 9 is twentieth in the ranking based on difficulty, and Item 33 occupies the rank of 19 on difficulty.

#### RESULTS

How do these percentages and these ranks compare with those observed in European studies? Ferracuti and Rizzo (1959) published item percentages for a sample of 200 university students in Italy, tested under power conditions. The correlation between these percentages and those offered in Table 1 was  $+ .83$ . Considering the difference in samples and testing conditions, this would appear to be a rather high degree of consistency in the ranking of items by difficulty.

Rankings for the items based on a sample of 78 students, ages 11:6 to 13:6, were given by Vautrin (1954). The rank-order correlation between Vautrin's list and that in Table 1 was  $+ .89$ .

Finally, Cloet (cited by Pasquasy & Doutrepoint, 1956) classified the 44 items into three levels of difficulty on the basis of testing in a sample of 285 Flemish children. Arbitrary scores of 3, 2, and 1 were assigned

<sup>2</sup> We are indebted to Robert Barney for his kindness in authorizing the testing, and for his help in assembling the scholastic records and other test scores.

<sup>1</sup> Edgar Anstey, personal communication, November 17, 1961.



TABLE 1  
PERCENTAGE OF STUDENTS ( $N = 86$ ) ANSWERING EACH ITEM CORRECTLY ON THE D 48 TEST, AND  
RANK ORDER PLACEMENTS FOR EACH ITEM

Item number	% answer- ing cor- rectly	Rank order	Item number	% answer- ing cor- rectly	Rank order	Item number	% answer- ing cor- rectly	Rank order
1.	97	1	16.	74	11	31.	19	29
2.	97	1	17.	16	30	32.	14	31
3.	94	5	18.	14	31	33.	51	19
4.	95	3	19.	65	16	34.	8	35
5.	86	7	20.	63	17	35.	5	37
6.	95	3	21.	49	21	36.	33	26
7.	70	13	22.	7	36	37.	27	27
8.	70	13	23.	35	25	38.	2	40
9.	50	20	24.	10	33	39.	3	39
10.	85	8	25.	9	34	40.	2	40
11.	76	10	26.	22	28	41.	0	42
12.	84	9	27.	42	23	42.	5	37
13.	91	6	28.	48	22	43.	0	42
14.	72	12	29.	42	23	44.	0	42
15.	70	13	30.	52	18			

to Cloet's three categories, and then these scores were correlated with the percentages in Table 1, giving a coefficient of  $+ .91$ .

It may therefore be concluded that the relative difficulty level of the 44 items is rather constant for different age groups, and for testing in different countries and in different languages.

The comparison of average scores for a number of samples is found in Table 2. Data from Cusin (1959) show highest scores for engineering college graduates in Italy, proceeding through science graduates, economics and commerce, and lowest values for graduates in jurisprudence, letters, and philosophy. Cusin interpreted this progression as evidence for the validity of the D 48.

Maury's (1954) study in France shows a steady upward progression of D 48 scores for both sexes as samples are drawn from successively higher levels of the educational system. In samples of Belgian and Swiss school children, Pasquasy and Doutrepoint (1956) demonstrated a similar hierarchy of D 48 means for successively older groups. Our data, although deriving from only two classes, agree in showing a higher average score for sixth graders than for fifth grade students.

The data in Table 2 may also be reviewed

for comparison of absolute scores in different countries. Our fifth and sixth grade subjects showed modal ages of 10 and 11. Comparable European samples are the 11- to 13-year-old students of Vautrin, and the 10:6- to 11:5-year-old sample of Pasquasy and Doutrepoint. Our subjects had means of 18.68 and 20.02; Vautrin's subjects had a mean of 20.76, and those of Pasquasy and Doutrepoint a mean of 20.00. The correspondence here is close enough to suggest that average scores for subjects of similar age and educational level in different countries will be about the same.

A word might also be added about the sex differences indicated in Table 2. The most thorough consideration to date of this issue is that offered by Pasquasy and Doutrepoint, who concluded (p. 31) that sex differences were not significant and could therefore be disregarded in providing norms on the test. In the four instances in Table 2 where sex differences are shown, males score higher three times and females once. Our fifth grade sample ( $N = 34$ ) included 15 boys and 19 girls, and the 52 sixth graders included 28 boys and 24 girls. In both grades the girls had a slightly higher mean score than the boys. If these findings are added to those in

Table 2, the differences would favor males three times and females three times; the proper conclusion would appear to be that drawn by Pasquasy and Doutrepoint. However, this question should not be considered

closed, as more data and studies of larger samples are needed before a definitive answer can be proposed.

The predictive validity for the D 48 test in the fifth grade sample is shown in Table 3.

TABLE 2  
ILLUSTRATIVE NORMS ON THE D 48 TEST

Samples	N	M	SD
1. From Cusin (1959), Italy			
a. Graduates in engineering	160	32.46	4.67
b. Graduates in science	28	30.71	5.10
c. Graduates in economics and commerce	47	30.04	4.79
d. Graduates in jurisprudence, letters, and philosophy	28	27.96	5.98
2. From Ferracuti and Rizzo (1959), Italy			
a. University students, male	100	26.70	5.60
b. University students, female	100	24.80	6.20
3. From Maury (1954), France			
a. Primary school graduates			
1. Males	307	14.25	6.06
2. Females	246	13.15	6.82
b. Students, quatrième classe (ninth grade)			
1. Males	144	22.87	5.40
2. Females	191	23.73	6.94
c. Students, deuxième classe (eleventh grade)			
1. Males	60	27.70	4.68
d. Students, Baccalauréat (twelfth grade)			
1. Males	73	30.70	5.04
2. Females	118	27.82	5.12
4. From Pasquasy and Doutrepoint (1956), France			
a. Young men, ages 20-25	522	19.78	8.61
b. Engineers and military officers, males	118	30.34	5.11
5. From Pasquasy and Doutrepoint (1956), Belgium and Switzerland			
a. Students, male and female, ages 10:6 to 11:5.	34	20.00	5.75
b. Students, both sexes, ages 11:6 to 12:5	127	19.07	6.34
c. Students, both sexes, ages 12:6 to 13:5	199	19.33	6.45
d. Students, both sexes, ages 13:6 to 14:5	253	19.93	6.62
e. Students, both sexes, ages 14:6 to 15:5	181	20.21	7.14
f. Students, both sexes, ages 15:6 to 16:5	124	22.84	7.05
g. Students, both sexes, age 16:6 and over	87	25.57	6.72
6. From Vautrin (1954), France			
a. Students, both sexes, ages 11 to 13	78	20.76	5.36
7. Present study, United States			
a. Fifth grade students, both sexes	34	18.68	5.38
b. Sixth grade students, both sexes	52	20.02	5.86

TABLE 3  
CORRELATION MATRIX AMONG THE VARIABLES INDICATED,  
FOR A FIFTH GRADE CLASS ( $N = 34$ )

Variables	D 48	Arithmetic	Language	Reading	Spelling
1. Grade point average	.58	.42	.52	.51	.26
2. D 48 test	—	.43	.43	.51	.25
3. Stanford Achievement Test: arithmetic		—	.51	.54	.05
4. Stanford Achievement Test: language			—	.42	.02
5. Stanford Achievement Test: reading				—	.18
6. Stanford Achievement Test: spelling					—

The D 48 test was given in the fall, as were the four Stanford Achievement Test subtests. Grades covering 12 aspects of school performance, viz., arithmetic, art, citizenship, music, oral language, physical education, reading, science, social studies, spelling and writing, and written language, were obtained the following June, and an average computed for each child.

The D 48 score correlated +.58 with this average, a higher value than those observed for any of the four parts of the Stanford Achievement Test. If the "nonacademic" subjects of art, citizenship, music, and physical education are removed from the grade point average the correlations are essentially unchanged. It should also be noted that D 48 correlates more highly with grades than with any of the four parts of the Stanford Achievement Test. Grades from the previous grade

(fourth grade) were available for these 34 students. The correlation between the fourth and fifth grade GPAs was +.60, only slightly higher than the +.58 between D 48 and fifth grade GPA.

Similar data for the sixth grade class are offered in Table 4. Once again, testing was done in the fall, and then the grades in the same 12 categories listed for the fifth graders were collected in the following June, and averaged. The predictive validity of the D 48 test was +.45, again higher than the values observed for the achievement test.

If the GPA for sixth grade is based on only the eight "academic" subjects, the correlation with the D 48 test remains at +.45, but drops for all three scores from the Stanford Achievement Test. The GPA from the previous year (fifth grade) was again available, and this prior GPA correlated +.41

TABLE 4  
CORRELATION MATRIX AMONG THE VARIABLES INDICATED,  
FOR A SIXTH GRADE CLASS ( $N = 52$ )

Variables	D 48	Comprehension	Reasoning	Total
1. Grade point average	.45	.25	.37	.35
2. D 48 test	—	.27	.51	.45
3. Stanford Achievement Test:				
a. Arithmetic comprehension		—	.61	.86
b. Arithmetic reasoning			—	.93
c. Total score				—



with the sixth grade GPA; for this sample the D 48 test was a better predictor of sixth grade GPA than were fifth grade GPAs. The conclusion appears justified that the D 48 is an adequate predictor of scholastic performance at the grade school level.

#### SUMMARY

The D 48 test of general ability is widely known and used in Europe, but is little known in the United States and is unmentioned in standard references on psychological testing. This state of affairs is unfortunate, and should be rectified, as the test has a number of very attractive features, including these: it is highly saturated with *g*, it is easy to give and to score, it is almost entirely nonverbal, and it possesses an intrinsic potentiality for cross-cultural usage.

It was therefore deemed appropriate to undertake a study of the D 48 in this country, with the intent of evaluating (a) item difficulties, (b) comparability of American and European norms, and (c) the predictive validity of the instrument.

For this investigation, 86 grade school children were tested. Item difficulties were calculated and were found to correlate very highly with similar data from European samples. Means and standard deviations for the American samples were likewise very close to the values observed for similar samples tested in Europe. Finally, the predictive validity for grade averages in fifth and sixth grades were  $+ .58$  and  $+ .45$ , respectively, both coefficients being higher than any observed for the subtests of the Standard Achievement Test on these same samples.

It was concluded that the D 48 test is applicable to school children, that its results

are meaningful and interpretable according to norms generally available, and that the test possesses adequate predictive validity. It is hoped that these results and the attendant discussion will prompt other investigators to initiate additional studies of the D 48. The test would appear to merit a thorough and widespread consideration by psychologists in the United States.

#### REFERENCES

- ANASTASI, ANNE. *Psychological testing*. (2nd ed.) New York: Macmillan, 1961.
- BUROS, O. K. (Ed.) *The fourth mental measurements yearbook*. Highland Park, N. J.: Gryphon Press, 1953.
- BUROS, O. K. (Ed.) *The fifth mental measurements yearbook*. Highland Park, N. J.: Gryphon Press, 1959.
- BUROS, O. K. (Ed.) *Tests in print*. Highland Park, N. J.: Gryphon Press, 1961.
- CALONGHI, L. *Testi e esperimenti*. Turin, Italy: Ponteficio Ateneo Salesiano, 1956.
- CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper, 1960.
- CUSIN, S. G. Contributo alla taratura italiana del D 48. *Boll. Psicol. Sociol. appl.*, 1959, No. 31-36, 259-261.
- FERRACUTI, F., & RIZZO, G. B. Studio sul test D-48 applicato ad una popolazione italiana di livello scolastico superiore. *Boll. Psicol. Sociol. appl.*, 1959, No. 31-36, 77-83.
- MAURY, H. Etude des résultats obtenus à l'aide du D48 à différents niveaux. *BINOP*, 1954, 10, 165-177.
- PASQUASY, R., & DOUTREPONT, G. Le test des dominos (D48). *Bull. Orient. scol. profess.*, 1956, 5, 20-34.
- PICHOT, P. *Les tests mentaux en psychiatrie*. Paris: Presses Universitaires France, 1949.
- PICHOT, P., RENNES, P., & TAVER, G. Quelques considérations à propos des tests D48 et vocabulaire. *Rev. Psychol. Appl.*, 1953, 4, 395-405.
- VAUTRIN, H. Etude du D48. *BINOP*, 1954, 10, 158-164.
- VERNON, P. E. *The structure of human abilities*. London: Methuen, 1950.

(Received December 20, 1962)

## SCALING THE SCALES: USE OF EXPERT JUDGMENT IN IMPROVING THE VALIDITY OF QUESTIONNAIRE SCALES<sup>1</sup>

K. WARNER SCHAIE

*University of Nebraska*

It is hypothesized that the concurrent validity of questionnaires can be increased by the use of item weights obtained by expert scaling instead of using the conventional unit weights. Items from 11 scales of the High School Personality Quiz (HSPQ) were rewritten in a form permitting application of the constant sum method to the judgment of item weights. The scales were rated by 30 psychologists and ratio scales constructed for each factor scale. The HSPQ records of 43 delinquent girls were correlated with behavior ratings on 42 traits from Cattell's "normal trait sphere" as judged by the Ss' cottage parents. Validity coefficients are given for the HSPQ unit weight scores and for scores weighted by expert-derived ratio weights. Results show low magnitude increments in validity. Effects of scaling on the problem of agreement among questionnaire and rating data were also evaluated with similar results.

The traditional procedure in constructing questionnaire scales consists of defining an item pool and using empirical methods to derive subsets of items which are relevant to some observed or postulated dimension. Unit weights are usually assigned to all items because there is no information on relative item weights. Even when factor weights offering greater precision are available they are generally ignored for practical reasons or because it is argued that increments in precision due to differential item weights will be trivial.

The items selected for a given questionnaire scale on empirical grounds may in fact be the best items available but there is no assurance that the resulting unit weight scale has certain properties which are likely to maximize validity. An important property for this purpose would be the requirement that the questionnaire scale be a ratio scale. Factor analytic or other empirical procedures of scale construction can obviously provide no assurance of this property.

<sup>1</sup> A preliminary version of this paper was presented at the annual meeting of the Society for Multivariate Psychology, November 1962, at Chicago, Illinois. The author is grateful for the cooperation of his colleagues and graduate students who performed the expert scaling, and to the staff at the State Training School for Girls, Geneva, Nebraska, for their cooperation in the testing and behavior rating of the questionnaire respondents.

It may furthermore be argued that the ultimate criterion of the utility of a questionnaire will be found in its ability to predict traits or behaviors which are confirmable by direct observation. The psychologist in his role as an expert on human behavior may have relevant information on the relations between self-report and observable traits. It is therefore proposed that an attempt be made to utilize the intuitive processes and capitalize upon the training and sophistication of the psychologist by utilizing his judgment to improve the weights assigned to data derived from actuarial methods. As has been pointed out by Meehl (1954), this does *not* mean that the psychologist is a superior computer. It means that he has been programmed with rules which are not yet explicit and which are therefore not available for direct implementation. Such rules would include that of rejecting data which are patently absurd although possible, or of assigning different weights or discounting weights for data which for computational efficiency have been handled as if they were of equal weight even though no such equality exists.

In spite of the quite popular feeling that there is a wide gulf between psychometric and psychophysical procedures, one may utilize the latter to advantage to maximize



desirable scale properties. The constant sum method (Metfessel, 1947) provides a promising approach for obtaining ratio scales on more or less subjective psychological variables. The method has been applied mostly where physical standards were also available, but it has been found useful and reliable in experiments scaling such matters as preference for neckties (Dudek & Baker, 1956) and the association between colors and mood tones (Schaie, 1961).

It is now proposed to apply this constant sum method to the scaling of empirically derived personality scale items and to examine the effect of such scaling on the concurrent validity of the questionnaire scales in predicting trait ratings and factor scores based on such trait ratings.

#### PROCEDURE

The High School Personality Quiz (HSPQ) was developed by Cattell, Beloff, and Coan (1958) to provide a comprehensive personality description of adolescents by means of factored scales. The HSPQ contains 14 questionnaire scales each consisting of 10 items with unit weights. According to the authors, 11 of the 14 scales assess factors which have also been identified in the behavior rating domain. This study is therefore limited to the 11 scales for which validity data are available.

The items on the 11 HSPQ scales on Form A of the questionnaire (Factors, A, B, C, D, E, F, G, H, I, J, and O) were examined to determine whether they were in unidirectional form. The items which did not meet this criterion were then rewritten. For example, the item "At a picnic, would you rather spend some time: a. exploring the woods alone? b. playing around the campfire with the crowd?" was rewritten in the form "At a picnic I would rather spend some time playing around the campfire with the crowd than exploring the woods alone."

The rephrased items were arranged in a booklet with the 10 items of each scale on one page, appearing in the order in which the items appear in the questionnaire. These booklets were then handed to 30 psychologists (their qualifications being defined here as eligibility for election as APA associates) who were asked to scale items by means of the constant sum method.

The constant sum method requires the judge to indicate the relative magnitude of two stimuli by dividing 100 points between them. Although the procedure generally used in the constant sum method is patterned after the method of paired comparison there is no need for this particular model. In fact the large number of indirect estimates may introduce different judgmental biases depending on the dissimilarity of the stimuli which are to be com-

pared. For this reason, as well as to keep the judgmental task within reasonable bounds, a modified technique using the method of the constant stimulus as a model was applied. This modification was originally applied to the scaling of line lengths (Baker & Dudek, 1957), where physical dimensions are available, but has also been found useful and reliable in the scaling of the association of colors and mood tones (Schaie, 1961).

The task of the expert judges was to compare Item 1 on each scale with each of the other nine items and to divide 100 points between each item pair assigning the larger number to the item rated more appropriate to the positive end of the factor dimension. As an example the instructions for Factor A (Cyclothymia versus Schizothymia) were as follows:

Compare statement 1 with each of the following statements. Divide 100 points between each pair of statements to judge their relative position on the dimension "warm, sociable vs. stiff, aloof." Record only the number of points assigned to statement 1. The larger number of points should be assigned to the statement judged to be more "warm, sociable."

Scale values were obtained by totalling the points assigned to Item 1 in comparison with all the other items and the ratio of the sums of all possible points to the total for Item 1 was calculated for each of the compared items. To permit comparison with the unit weights of the original questionnaire scale values were next transformed to a 10-point scale. On practically all scales some item was rated as half as relevant or important as some other item with respect to the dimension on which it had been placed. For example, the item weights obtained for the scale for Factor A were 1.4, 1.0, 0.9, 0.9, 0.8, 0.7, 1.6, 1.1, 0.7, 0.9.

The new item weights were next applied to the rescaling of a set of HSPQ protocols obtained from a group of 43 girls who were residents of a training school for delinquent adolescents. These girls were also rated by their cottage parents on the 42 traits described by Cattell (1957) as providing an adequate sample of what he calls the "normal trait sphere." Furthermore, factor scores were computed from these trait ratings using Cattell's weights which yielded measures for 15 rating factors. Intercorrelations of the HSPQ measures, validity coefficients describing the relationship between the HSPQ and the trait ratings and rating factor scores, as well as multiple correlations between the questionnaire and criterion data, were then obtained to compare the effect of the new scores using the weights provided by the experts as contrasted to the original unit weights. The following section will discuss the results of these analyses in some detail.<sup>2</sup>

<sup>2</sup> Tables giving the ratio item weights for each of the 11 scales, means, standard deviations, scale intercorrelations, and multiple correlations with the criterion variables have been deposited with the American Documentation Institute. Order Document No. 7557 from ADI Auxiliary Publications Project,



TABLE 1

MEAN DIFFERENCES, CORRELATIONS, AND PROPORTION OF INDEPENDENT VARIANCE  
BETWEEN UNIT WEIGHTED AND RATIO WEIGHTED SCORES

	Factors										
	A	B	C	D	E	F	G	H	I	J	O
Mean difference	-.02	.08	.26	-.18	.18	.05	-.07	-.13	-.14	-.18	-.24
Correlation	.97	.95	.94	.96	.96	.97	.98	.99	.98	.95	.98
Proportion of independent variance	6%	10%	12%	8%	8%	6%	4%	2%	4%	10%	4%

### RESULTS

The first question to be examined is the effect of the scaling upon the descriptive statistics of the HSPQ performance of our sample of questionnaire respondents. In evaluating differences between means the proper alternative to the null hypothesis is that the means for the scaled scores should be lower than the means for unit weight scores since unit weight scores should tend to overestimate scores of individuals rating low on a given scale. A difference significant at the .05 level of confidence appears only for Factor O. However, keeping in mind the relatively small sample and the accompanying danger of Type II error, it should be reported that differences in the expected direction which were significant at the .10 level of confidence were found also for Factors D, H, I, and J. An inspection of score variability further indicates that variability increases as a result of scaling for all scales except Factors H and I.

Intercorrelations were also computed for the two score matrices and the Kolmogorov-Smirnow test of the difference between two correlation matrices (Schaie, 1958) was applied. It was found that the two matrices did not show more than chance deviation from one another and that the scaling therefore apparently had no effect upon the general structure of scale intercorrelations. This is desirable since the factor-analytic solution presumably has obtained the best set of

items with respect to factorial structure and optimal estimation of factor scores.

Table 1 lists the mean differences between unit weight and ratio scaled scores, as well as the correlations and the proportions of independent variance available for each scale. The latter, ranging from 2% to 12%, represent the limit of the increment in precision which may be expected as a result of the scaling procedure. The magnitude of this increment is small, but it may be artificially restricted by requiring adherence to the original 10-point scale.

Our basic premise for this study was the hypothesis that the concurrent validity, i.e., the statements which can be made about the ability of the HSPQ to predict criterion behavior, would be increased by expert weighting of items. The next logical step therefore is an examination of changes in the validity coefficients. These are reported in Table 2. It may first of all be noted that the number of validity coefficients which may be interpreted as differing significantly from zero at the .05 level of confidence has increased after scaling. Sixty-eight coefficients significant by the above criterion were found for the unit-weight scores, while 77 such coefficients appeared after applying the scaled weights. Of the 81 coefficients reaching the .05 level of significance by either scoring method, 45 showed increase in magnitude after scaling, 18 retained the same magnitude, while 18 showed some decrease. No changes in magnitude, however, exceed 1 point in the first decimal place.

Questionnaire results are rarely interpreted scale by scale. More frequently use is made of the scale profile or regression equations

TABLE 2  
BISERIAL CORRELATIONS BETWEEN HSPQ FORM A AND TRAIT RATINGS

Traits	HSPQ Factors										
	A	B	C	D	E	F	G	H	I	J	O
1. Considerate-inconsiderate	-.11 -.06	.36* .41*	.16 .14	-.24 -.17	.12 .12	-.05 -.03	.15 .20	-.07 -.06	-.20 -.18	.06 .16	-.08 -.12
2. Calm-excitable	-.20 -.19	-.12 -.17	-.23 -.22	-.16 -.16	.05 .03	-.19 -.25	-.17 -.16	-.02 -.05	.09 .10	.38* .41*	-.04 -.12
3. Energetic-tired	.08 .06	.11 .13	.10 .07	-.06 -.12	.00 -.04	.34* .33*	-.06 .00	.25 .24	-.25 -.21	-.21 -.25	-.23 -.20
4. Quiet-noisy	-.03 -.02	.13 .12	-.42* -.36*	.19 .16	.03 -.11	-.43 -.49*	.14 .13	-.39* -.38*	.13 .17	.28 .34*	.08 .05
5. Patient-impatient	.16 .13	.22 .21	.00 .04	-.16 -.16	.10 .02	.00 -.05	.19 .22	.05 .08	-.28 -.25	.06 .10	-.25 -.27
6. Cheerful-solemn	.01 .02	-.04 -.05	.04 -.01	-.12 -.13	.16 .27	.41* .40*	-.24 -.28	.23 .21	-.19 -.22	-.25 -.30	-.07 .00
7. Friendly-reserved	-.03 -.04	-.28 -.20	.32* .35*	-.24 -.23	.00 -.09	.55* .50*	.03 .06	.17 .16	-.14 -.17	-.40* -.42*	-.25 -.24
8. Meditative-unquestioning	.22 .22	.29 .24	.03 .04	-.09 -.16	-.24 -.31*	-.30* -.33*	.49* .45*	-.07 -.01	.17 .16	.11 .22	-.25 -.31*
9. Cooperative-obstructive	-.10 -.03	.23 .28	.40* .44	-.03 -.02	.09 .12	.24 .25	.36* .41*	.05 .03	-.27 -.27	-.09 -.04	-.18 -.22
10. Happy-sad	-.23 -.24	.02 .06	.00 -.04	-.31* -.36*	.13 .11	.23 .25	-.04 -.01	.08 .09	-.06 -.07	.03 .00	-.10 -.09
11. Sensitive-tough	.06 .15	.33* .32*	.12 .21	-.11 -.04	.13 .05	-.10 -.15	.40* .44*	-.08 -.09	-.23 -.21	.26 .32*	-.12 -.15*
12. Intelligent-stupid	-.02 .01	.33* .26	-.07 -.06	.42* .46*	-.01 -.02	.04 .04	.31* .38*	-.05 -.04	-.29 -.26	-.29 -.28	.09 .04
13. Poised-flustered	-.26 -.35*	.05 .02	-.25 -.24	.04 -.01	-.02 .00	-.21 -.21	-.12 -.10	.03 -.01	.16 .21	.10 .10	.03 -.02
14. Tolerant-jealous	.17 .19	.04 .07	.08 .22	.00 .08	.15 .13	-.07 -.10	.21 .23	.22 .20	-.28 -.26	.22 .26	-.28 -.31*
15. Dominant-submissive	-.10 -.12	-.04 .00	.00 -.07	-.07 -.06	-.29 -.19	.26 .27	-.25 -.30*	.00 .00	.17 .13	-.46* -.44*	.09 .12
16. Relaxed-tense	-.12 -.09	-.11 -.09	-.35* -.25	-.18 -.12	.08 .00	.01 -.01	-.05 .02	.09 .05	.21 .21	.17 .22	.08 .05
17. Conventional-unconventional	-.16 -.12	.22 .19	.19 .26	.17 .18	.04 .00	-.21 -.20	.25 .26	.05 .05	.17 .19	.14 .17	-.15 -.16
18. Sociable-self-contained	.10 .13	-.50* -.46*	.07 .03	-.19 -.17	.23 .26	.30* .27	-.16 -.17	-.11 -.15	-.13 -.17	-.03 -.02	.29 .31*
19. Trustful-suspicious	-.10 -.12	.20 .18	.19 .32	.11 .14	.15 .19	-.07 -.05	-.13 -.12	.16 .18	-.17 -.14	.02 .09	-.06 -.10
20. Self-effacing-egotistical	.16 .14	.16 .12	.11 .18	.10 .04	.12 .08	-.04 -.06	-.03 -.03	-.17 -.18	-.04 .01	.19 .30*	.01 .00
21. Conscientious-unscrupulous	.18 .21	-.18 -.19	.16 .21	-.08 -.15	.24 .22	.04 .03	.26 .24	-.22 -.22	-.18 -.18	.09 .14	-.16 -.20

TABLE 2—Continued

Traits	HSPQ Factors											O
	A	B	C	D	E	F	G	H	I	J		
22. Adventurous-timid	12 15	-12 -12	48* 46*	-10 -02	-12 00	41* 43*	14 15	54* 54*	-38* -41*	-34* -45*	-30* -29	
23. Stable-unstable	-02 -07	34* 34*	-08 -02	-03 -04	-03 -05	-06 -01	28 37*	13 13	-29 -25	-26 -19	04 -02	
24. Persevering-quitting	09 13	04 03	01 04	-34* -39*	-21 -28	01 -01	28 30*	03 05	-02 -05	06 12	-17 -21	
25. Modest-attention-seeking	-22 -21	32* 34*	-27 -31	-02 -09	03 -04	-30* -26	08 09	-43 -42*	17 20	32 40	19 16	
26. Open-defensive	23 26	-01 03	-15 -12	03 -01	26 26	-03 -01	-16 -10	-48* -49*	-36* -34*	33* 37*	38* 32*	
27. Refined-crude	-18 -17	15 06	-24 -21	08 -01	29 23	-20 -22	-05 -03	-42* -40*	-11 -06	06 17	23 21	
28. Imaginative-practical	-01 -07	-39* -44*	43* 48*	00 05	-13 -07	06 01	-09 -17	30* 33*	-07 -06	-07 -15	-37* -32	
29. Obedient-disobedient	-01 01	08 04	38* 46*	-27 -28	03 07	02 -02	15 12	02 05	-30* -29	-07 01	-30* -34*	
30. Adaptable-inflexible	-19 -17	35* 34*	11 20	-11 -09	15 21	-21 -17	18 21	16 15	-33* -28	-01 04	-12 -15	
31. Responsible-irresponsible	09 16	20 16	03 06	-15 -18	-05 -06	-10 -12	59* 61*	-20 -16	-16 -19	-05 04	-12 -15	
32. Curious-incurious	-07 -09	-17 -12	47* 37*	00 -03	-04 09	30* 36*	06 02	17 17	-21 -20	-41* -42*	-22 -02	
33. Talkative-silent	09 08	03 01	15 11	06 09	06 24	37* 39*	-07 -09	16 16	-36* -37*	-36* -48*	03 -18	
34. Carefree-anxious	-06 -09	-15 -13	30* 32*	05 05	-06 06	44* 48*	-34* -33*	25 25	-16 -20	08 -01	-16 -17	
35. Tasteful-inartistic	18 22	-24 -28	20 21	-02 -03	-09 -07	27 20	18 16	-08 -06	-05 -07	-20 -12	-18 23	
36. Resourceful-baffled	-15 -17	-11 -11	-12 -09	39* 44*	23 23	-02 05	-05 03	-02 -01	-42* -42*	-09 -13	18 -04	
37. Independent-dependent	15 17	09 07	08 12	-02 01	08 09	07 08	08 12	09 12	-37* -38*	00 00	-08 08	
38. Adult-naive	03 00	17 14	03 04	33* 27	-10 -07	07 13	28 35*	13 13	-29 -22	-39* -35*	03 18	
39. Orderly-disorderly	-06 05	02 -05	-23 -21	-29 -30*	-01 -08	-33* -35*	15 12	-30* -29	-02 -04	15 25	14 -12	
40. Easygoing-irritable	-15 -08	-15 -09	39* 35*	-25 -16	24 28	00 01	-04 00	02 01	-45* -51*	15 16	-18 -03	
41. Expressive-secretive	03 01	-23 -20	04 08	-16 -15	10 11	16 12	05 04	-06 -05	01 02	-14 -13	-03 -25	
42. Brave-complaining	23 29	-04 -02	27 32*	05 01	10 08	20 18	19 18	-06 -06	-08 -08	11 16	-28	

Note.—The top figure in each row is for the unit weighted scores, the bottom figure is for the ratio weighted scores. In bold face indicate correlations with the factor on which a given trait is hypothesized to load (Cattell, 1957).

\* $p < .05$ .

Figures



based on known relations between the scales and criterion measures. It was therefore felt to be appropriate to compute multiple correlations between the HSPQ and each of the 42 trait rating criterion variables. To avoid maximizing chance and artifactual relationships in the sample use was made of a technique described by Horst and Smith (1950) as programed for the Burroughs 205 computer by Schaie (1963). This is an iterative method which selects the largest validity coefficient as the first estimate of

$R$ , subtracts out the product of this value with the appropriate vector of the predictor matrix, and then selects the next largest remaining validity coefficient to be added. After each cycle the multiple  $R$  is corrected for shrinkage by means of Wherry's formula. Whenever the corrected  $R$  no longer increases or when it decreases, then the multiple  $R$  obtained at the end of the previous cycle is accepted as the best estimate of  $R$  and further iteration is ceased.

Of the 42 regression equations computed,

TABLE 3

CORRELATIONS BETWEEN HSPQ FORM A AND FACTOR SCORES DERIVED FROM TRAIT RATINGS

Trait Rating Factors	HSPQ Factors										
	A	B	C	D	E	F	G	H	I	J	O
A Cyclothymia versus schizothymia	-.07 -.06	-.02 -.01	.15 .21	-.17 -.14	.17 .22	.06 .05	-.08 -.06	.03 .02	-.24 -.23	.05 .09	-.03 -.06
B Intelligence	.06 .09	.22 .20	-.12 -.10	.03 .00	-.22 -.23	-.03 -.03	.38* .40*	-.11 -.09	.02 .02	-.13 -.08	-.07 -.10
C Ego strength versus proneness to neuroticism	.07 .07	.00 -.02	-.04 .02	-.10 -.13	.06 .04	.00 -.01	.09 .12	.00 .00	-.11 -.10	.15 .18	-.12 -.17
D Excitability versus insecurity	-.07 -.09	-.06 -.06	-.04 -.09	.14 .16	-.05 -.01	.04 .05	-.11 -.11	.04 .03	.08 .07	-.25 -.31*	.15 .18
E Dominance versus submissiveness	-.01 -.02	-.28 -.26	.05 .01	.03 .04	-.08 -.02	.23 .25	-.15 -.16	.06 .05	.01 -.02	-.23 -.29	.12 .13
F Surgency versus desurgency	-.12 -.12	.11 .08	.16 .19	-.08 -.06	.17 .24	.16 .19	-.08 -.05	.28 .27	-.27 -.25	-.07 -.10	-.13 -.14
G Super ego strength	.02 .08	.11 .07	-.04 -.01	-.21 -.22	.03 .00	-.13 -.15	.26 .27	-.14 -.12	-.08 -.09	.14 .22	-.08 -.12
H Parmia versus threectia	-.06 -.06	-.23 -.17	.39* .36*	-.14 -.09	.06 .19	.41* .43*	-.06 -.04	.22 .21	-.31* -.35*	-.26 -.33*	-.07 -.07
I Premsia versus harria	-.19 -.20	-.08 -.12	.13 .12	.03 .07	-.12 -.10	-.14 -.16	.05 .00	.08 .06	.25 .26	-.02 -.01	-.03 -.01
J Coasthenia	.01 .02	-.04 -.06	-.29 -.29	-.14 -.12	-.11 -.17	-.44* -.47*	.06 .02	-.17 -.15	.33* .32*	.46* .49*	.10 .08
K Contention versus abcultion	.10 .11	-.11 -.14	.08 .10	-.05 -.06	.09 .10	.12 .10	.09 .08	-.11 -.10	-.12 -.11	-.04 -.02	-.15 -.17
L Protension versus inner relaxation	.02 -.01	-.04 -.08	-.30* -.39*	.10 .06	-.09 -.13	-.19 -.18	-.12 -.14	-.19 -.18	.18 .18	.03 -.01	.24 .26
M Autia versus Praxernia	-.04 -.08	-.20 -.24	-.14 -.12	-.01 -.01	-.15 -.12	.05 .03	-.04 -.03	.07 .07	.00 -.01	-.05 -.06	.03 .00
N Shrewdness versus naivete	.14 .17	.17 .10	-.08 -.09	.24 .22	.20 .19	.00 .01	.09 .13	-.23 -.23	-.25 -.21	.01 .08	.14 .10
O Guilt proneness versus confidence	-.01 .00	-.03 -.01	.05 -.01	-.12 -.10	-.25 -.24	-.18 -.19	.11 .08	-.15 -.14	.03 -.01	.03 .02	.14 .15

Note.—The top figure in each row is for the unit weighted HSPQ scores, the bottom figure is for the scores ratio weighted by the Constant Sum Method. Figures in bold face indicate correlations between matching factors in the behavior rating and questionnaire domains.

\*  $p < .05$ .

28 resulted in increased and 14 in decreased corrected multiple  $R$ s. On the average, scaling raised multiple  $R$ s by .037 and corrected multiple  $R$ s by .052. Twenty of the multiple correlations based on ratio weighted scores showed an increase in prediction significant at or beyond the .05 level of confidence. Carrying out the iterative process until all predictors had been assigned a weight changed individual coefficients but did not affect overall results. The reported shifts are not large, but they do seem to indicate a noteworthy trend.

#### EFFECTS OF SCALING ON CROSS-MEDIUM FACTOR MATCHING

An interesting extension of our experiment leads to the question whether it can contribute some clarification to the controversy regarding the matching of L (behavior rating) and Q (questionnaire) data (Becker, 1960; Cattell, 1961). Specifically we are concerned with the question of what if any effect the scaling will have in reducing the effects of such variables as the instrument and perturbation factors specified by Cattell to account for the lack of congruence between L and Q data. A more thorough approach to this problem would require factor analytic studies, which for the unscaled data have been reported elsewhere (Schaike, 1962), but which were not attempted in the present context. We shall be concerned here primarily with an examination of the effects of scaling on the zero-order correlations. If instrument and perturbation factors in fact conceal similarities in factor structure, then the increment in validity obtained by expert scaling may be a function of the decrease of effects of these semiartifactual components.

To study these matters, validity coefficients were computed between both scaled and unscaled HSPQ scores and Rating Factor scores which were obtained by combining trait ratings according to the factor structure reported by Cattell (1957). These coefficients are reported in Table 3. Matching between factor scores from the R and Q media is obtained at the .05 level of significance only for Factor J (Coasthenia). Although none of the coefficients for Factor G (Ego strength)

reach significance, the matching loadings are highest. In terms of our significance criterion, mismatches occur between  $C_q$  and Factors  $H_1$  and  $L_1$ ; between Factor  $F_q$  and Factors  $H_1$  and  $J_1$ ; between Factor  $G_q$  and Factor  $B_1$ ; between Factor  $I_q$  and Factors  $H_1$  and  $J_1$ ; and between Factor  $J_q$  and Factors  $D_1$  and  $H_1$ . What is the effect of the scaling on these relationships? In our sample it is restricted to raising most of the significant coefficients very slightly and in introducing two additional mismatches.

#### CONCLUSIONS

It is apparent from these results that there seems to be a consistent trend for an increase in the concurrent validity of the HSPQ scales as well as their combination after rescaling with expert-derived ratio weights. It is also apparent that the magnitude of the increases for our sample is not very great and the legitimate question arises whether the increment is worth the extra labor involved. It should be kept in mind, however, that our sample of questionnaire respondent was quite limited and that the criterion variables were certainly fallible. Also, we do not know what effect the rewriting of the questionnaire items had in modifying the judges' decisions. The results with respect to the cross-medium factor matching question are further limited by the fact that we used Cattell's factor weights for adults in computing factor rating scores. The factor structure of rating data on adolescents may, however, be somewhat different than that for adults, resulting in possible erroneous conclusions from our data. It would seem desirable, therefore, to conduct further studies similar to the one described here under more favorable conditions, since it appears that expert judgment can indeed increase the validity of questionnaires. The task then remains to investigate the experimental conditions under which the gain from expert judgment will be sufficiently large to merit practical application.

#### REFERENCES

- BAKER, KATHERINE E., & DUDEK, F. J. Scaling line-lengths with modification of the constant-sum method. *Amer. J. Psychol.*, 1957, 70, 81-86.

- BECKER, W. C. The matching of behavior rating and questionnaire factors. *Psychol. Bull.*, 1960, 57, 201-212.
- CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
- CATTELL, R. B. Theory of situational, instrument, second order, and refraction factors in personality structure research. *Psychol. Bull.*, 1961, 58, 160-174.
- CATTELL, R. B., BELOFF, H., & COAN, R. W. *Handbook for the High School Personality Questionnaire*. Champaign, Ill.: Institute for Ability and Personality Testing, 1958.
- DUDEK, F. J., & BAKER, KATHERINE E. The constant-sum method applied to scaling subjective dimensions. *Amer. J. Psychol.*, 1956, 69, 616-624.
- HORST, P., & SMITH, S. The discrimination of two racial samples. *Psychometrika*, 1950, 15, 271-289.
- MEEHL, P. E. *Clinical versus statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954.
- METFESSEL, M. A proposal for quantitative reporting of comparative judgments. *J. Psychol.*, 1947, 24, 229-235.
- SCHAIK, K. W. Tests of hypotheses about differences between two intercorrelation matrices. *J. exp. Educ.*, 1958, 26, 241-245.
- SCHAIK, K. W. Scaling the association between colors and mood tones. *Amer. J. Psychol.*, 1961, 74, 266-273.
- SCHAIK, K. W. On the equivalence of questionnaire and rating data. *Psychol. Rep.*, 1962, 10, 521-522.
- SCHAIK, K. W. Multivariate prediction programs for the Burroughs 205 computer. *Behav. Sci.*, 1963, 8, 170-172.

(Received December 28, 1962)



## COUNSELOR ANXIETY IN RELATION TO AMOUNT OF CLINICAL EXPERIENCE AND QUALITY OF AFFECT DEMONSTRATED BY CLIENTS<sup>1</sup>

PETER D. RUSSELL

*Buffalo Veterans Administration Hospital*

AND WILLIAM U. SNYDER

*Ohio University*

A counseling paradigm employed 2 actors role playing as clients, and 20 counselors. Independent variables were hostile or friendly client behavior and amount of counselor experience. There were 4 measures of the dependent variable of counselor anxiety: palmar sweating, eyeblink rate, client-actor estimates of counselor anxiety, and independent judgments of verbal anxiety of counselors' protocols. Results revealed that hostile client behavior led to significantly greater anxiety than friendly behavior. Amount of graduate training and counseling experience had little effect on the degree of counselor anxiety in either hostile or friendly interviews. Modification of some of the measures, lengthening of interviews, increasing group sizes, or finding more discriminate groups might have changed this last finding. Palmar sweating was of questionable utility as a measure of anxiety.

In recent years increasing attention has been given to the behavior of the counselor in interview situations. Studies have been made investigating counselor personality, counselor techniques, and the relationship between counselor and client. These questions have practical significance for graduate programs in clinical psychology since they are relevant to the teaching of counseling methods.

In the present study the investigators attempted to examine the effect of a hostile or a friendly client demeanor upon counselor anxiety from the standpoint of relatively experienced and relatively inexperienced graduate student clinical psychology counselors. Two hypotheses were developed. First, it was proposed that student counselors, regardless of amount of experience, display greater anxiety in interviews with hostile clients than in interviews with friendly clients. The second hypothesis was that more

experienced student counselors display less anxiety than less experienced student counselors in interviews with both hostile and friendly clients. These hypotheses were examined by means of four measures of the dependent variable of anxiety, i.e., palmar sweating, eyeblink rate, client-actor estimates of counselor anxiety, and judgments of verbal anxiety made by independent judges.

A review of the literature revealed that this was a relatively new area of consideration. Although there has been some theorizing, only a few experimental studies were found which explored the effects of hostile client affect on counselor anxiety. There were some investigations which utilized physiological measures, but only one report was found in which the counselors' palmar-sweating index had been employed, and this was only tangentially related to the current inquiry (Light, 1951). Using actors in the roles of clients, as in the present study, had been undertaken only once before to the investigators' knowledge (Kounin, Polansky, Biddle, Coburn, & Fenn, 1956). However, for that study the actor did not take an active role in the evaluation nor was the behavior of the counselors under scrutiny. Also, studies were reported which were concerned with

<sup>1</sup> This study is based on a PhD dissertation conducted by Peter D. Russell at the Pennsylvania State University under the direction of William U. Snyder, Donald H. Ford, and Deno G. Thevaos. One of the tests used in the study was derived from scales devised by William U. Snyder, which had been based in part on previous work by Donald H. Ford and Oliver H. Bown.

speech patterns as an index of anxiety, but none of these inquired into counselor behavior. No investigations were found which directly related the counselor's level of experience to the anxiety produced in certain types of interviews.

Two studies have reported results relevant to those cited in the present investigation. Bandura, Lipsher, and Miller (1960), using analyses of actual therapy interviews (but not simulated interviews), found that therapists who typically expressed their own hostility in direct forms, and who displayed low need for approval, were more likely to permit and encourage their patients' hostility than were therapists who expressed little direct hostility and who showed high approval seeking behavior. Psychotherapists were more inclined to avoid hostility when it was directed toward themselves than when the patients directed their hostility toward others. The patients were more likely to drop the hostile topic, or to change the object of their hostility following therapists' avoidance reactions, than they were following the therapists' approach reactions.

Heller, Myers, and Kline (1963), in a subsequent study, copied in part after the present investigation, and employing client-actors in simulated therapy situations, have demonstrated that dominant client behavior will evoke dependent interviewer behavior, and dependent client behavior will evoke dominant interviewer behavior. They also showed that hostile client behavior will evoke hostile interviewer behavior, and friendly client behavior will evoke friendly interviewer behavior. They were not able to demonstrate, as was the present study, that hostile client behavior will evoke anxiety on the part of a therapist.

## METHOD

### *Design and Procedure*

The design of the study randomized the order in which counselors participated. Two client-actors were each seen by half of a group of 10 more experienced counselors in their first interviews, and by the other half of the group in the second. The same procedure was followed for 10 relatively inexperienced counselors. Moreover, each of the two client-actors was seen in a friendly role 20 times and in a hostile role an equal number of times.

The two client-actors were chosen from a group

of undergraduate males enrolled in a theatre arts course. It was felt that actors could assume the roles required most meaningfully. While having demonstrated acting proficiency, they had not appeared extensively in public roles where they might have been observed by the student clinicians. The actors were chosen on the basis of lack of neurotic trends, similarity of physical appearance and age, and ability to take the required roles and play them spontaneously in a hostile and in a friendly manner. They were paid for their work, in order to maintain their interest in the project.

The actors underwent 10 hours of training to respond in interviews in two different ways: in a very friendly manner which was positive and helpful, and in a hostile and negative manner. They were quickly able to fill in the missing elements of their "background" stories when this seemed called for in the simulated counseling situation. While the spontaneity of the situation was maintained, their ways of presenting problems, information, and affect became quite constant. In a pilot study which was carried out, it was discovered that the actors were presenting somewhat extreme characterizations of the typical friendly or hostile client roles, so that the counselors of the pilot study were able to detect that the situations were somewhat exaggerated, and therefore not real counseling situations. Consequently the roles were made less deviant from the mean level of affect, and in the actual experiment itself, it was ascertained that none of the counselor-subjects detected the true character of the study, nor realized that the clients were actors.

The utility of the counseling paradigm which made use of actors who took the roles of clients was explored at some length. It was felt that this design had unique advantages for the study of counselor anxiety. It offers a compromise between the reality, but lack of control, of the study of real counseling, and the fine control, but lack of reality, of other analogues to a counseling situation, i.e., films or tape recordings. The use of actors role playing in what seems to be a normal counseling situation to the subjects, permits the manipulation of an independent variable which is uncontaminated by the fact that the counselor knows he is being studied or by other extraneous factors which are difficult to control. The counselor not only thinks that the client is real, but also that he is helping him. For the present investigation each subject was able to act as his own control for the study of the effects of hostility and friendliness due to the counterbalanced design, so the need for matching was eliminated.

### *Dependent Measures*

*Palmar sweating.* Employing the revised method of Mowrer (1953), the fingerprint stain test of palmar sweating was administered before and after each interview following thorough handwashing. The densities of the stains were measured by means of an adaptation of Mowrer's apparatus which passes light through the stain onto a photoelectric



cell. The counselors were told that this measure was being used on them in order to determine congruence of counselor-client feeling, a measure which was being considered for predictive ability regarding the extent to which marginally motivated clients would apply themselves in the counseling process.

In order to take into account possible influences on the amount of sweating, several new controls were employed. At the outset, stains were made for both index fingers. Where the stains appeared equally dark, the right index finger was used for subsequent measurements. In the cases where one print was perceptibly darker than the other, the finger producing the darker stain was used. This controlled for the possibility of peripheral nerve damage to an extremity (Cross, Dodds, & Knights, 1960) and consequent potential reduced sensitivity to tension changes.

Secondly, the counselors were allowed to rest for at least 15 minutes after coming to the experimental room so that they might become accommodated to the situation. This time was used to give the instructions. Also approximately 15 minutes of rest was given between post- and pre-measures of the first and second interviews, respectively, in order to facilitate dissipation of any anxiety and concomitant physiological reactions. Temperature and relative humidity were measured continuously by means of hygrothermographs in both the counseling room and the monitor room where the palmar-sweating measurements were taken. Large changes in either might have affected the amount of sweating, but no significant variations were noted. Also, it seemed necessary to make sure that the counselors were not ill, since it is likely that fever might cause an increase in palmar sweating. In order to guarantee that there were no cases of fever or illness, this topic was brought up during the period before the first interview, and ruled out before the counseling began. Finally, hyperthyroidism can lead to an increase in palmar sweating (Walmer, 1961).<sup>2</sup> However this factor was adequately controlled by checking the pulse rate.

*Eyeblink rate.* By means of a concealed closed-circuit television system, blink rates of the counselors were observed and recorded during the second and fifth 5-minute portions of the half-hour interviews. A check on the accuracy of the investigator's counting of the number of blinks showed a very high degree of reliability of measurement.

*"Client's" Scale of Counselor Anxiety.* Following the completion of each interview while the counselor was having his postinterview stain test, the client-actor completed a scale rating the anxiety which the counselor had revealed in the interview. This scale consisted of 60 items selected from three sources: items modified from Snyder's (1961) *Client's Postinterview Attitudes Toward Therapy* scale and *Therapist's Personal Reaction Question-*

*naire*, and a group of 10 items prepared especially for the present study.

*Judgments of verbal anxiety.* This scale of criteria of verbal anxiety was developed from statements concerning counselor anxiety which appeared in the literature,<sup>3</sup> in addition to several criteria developed for the present study. These criteria were refined by being submitted to a group of five experienced and carefully trained judges who were doing full-time vocational counseling and psychotherapy in the Division of Counseling at the Pennsylvania State University. Those criteria which were adjudged "significant measures" by three of the five judges, and for which the judges agreed upon the appropriate examples submitted to them, were compiled into a scale of 14 items. However, this scale was further modified by being submitted to a group of six other experienced judges, who were the ones who later did the actual rating of the interviews. Two new criteria which appeared relatively frequently enough in the pilot-study interviews were added to the original scale, thus making a total of 16 criteria for judging counselor anxiety which could be observed in the interview situation or protocols.

The sixteen criteria of counselor anxiety were the following:<sup>4</sup> direct statement of anxiety; asking questions or changing the subject; \* interrupting; impersonal or premature interpretations; unnecessary reassurance; disapproval; intellectualizing, and not responding to client's feelings; introjection of references to the counselor's own experiences; \* unfinished sentences; \* repeating words or phrases; \* stuttering; \* blocking; poor voice quality, tremulousness, "cracking" voice, etc.; apologizing for some fault in counseling technique; inappropriate laughter; and other signs of anxiety.

Several of these signs of anxiety were not common among all counselors, but had a relatively high frequency of occurrence in the protocols of particular ones. The starred items were common in all interviews.

### *Establishing Interjudge Agreement*

The six judges spent approximately 12 hours in training as a group. During this period the meaning of each criterion and its application to sample interviews was discussed and clarified. It was required that an agreement of 88% be reached on each unit in the test sample before training was considered adequate.

Some question arises regarding the problem of controlling for interview identity with regard to knowledge on the part of the judge as to whether the interview was a friendly or hostile one. It would

<sup>3</sup> See studies by Dibner (1956), Dollard and Mowrer (1947), Kauffman and Raimy (1949), Lasswell (1935), Sanford (1942), Zimmerman (1950), Mahl (1956, 1959), Boomer and Goodrich (1961) and Krause (1961).

<sup>4</sup> The starred items are those signs of anxiety occurring more frequently and more uniformly.

<sup>2</sup> J. D. Walmer, personal communication, July 15, 1961.



be impossible, of course, to disguise this fact from the judges (who might also have guessed the identity of a few of the counselors). Some of the specific criteria of counselor anxiety which might be more influenced by this knowledge would be "direct statement of anxiety," "premature interpretations," "disapproval," and "apologizing for some fault in counselor technique." However, these were among the signs of counselor anxiety which occurred fairly infrequently, and therefore had relatively little effect on the score of counselor anxiety at which the judges arrived in each case.

### *Selection of Student Counselors*

Two groups of 10 male student clinicians were drawn from the total population of clinical psychology graduate students. The 22 eligible students were evaluated for clinical experience in order to place them in the "more experienced" or "less experienced" group. Using a modification of Tirnauer's (1959) questionnaire, 13 types of clinical experience were taken into consideration in arriving at composite "clinical experience" rankings. The groups, as finally constituted, were also evaluated for general ability (using the Miller's Analogy Test scores) and age, since these factors might lead to the expected results if the experienced counselors were basically more capable and/or older. Ability, as defined by MAT scores, was not found to be a factor differentiating the two groups; the experienced group was, however, something a little over a half year older than the inexperienced group. This difference was not considered to be significant.

### *Judgments of Client-Actor Affect*

The 40 interview typescripts were presented in an individually randomized sequence to three non-psychologist judges. They were asked to sort them into equal-sized piles in terms of client hostility or friendliness. The three judges performed this task without error, thus indicating that the client-actors had adopted clearly differentiable roles throughout the experiment.

## RESULTS

The hypothesis concerned with the differential effects of hostility and friendliness on both the experienced and inexperienced groups of counselors was analyzed by means of the Wilcoxon matched-pairs signed-ranks test, while the Mann-Whitney *U* test was applied to the data in comparing counselor anxiety in terms of the levels of counselor experience.

As far as the comparison of types of affect was concerned, it was found that eyeblink rate, the "Client's" Scale of Counselor Anxiety, and the scores derived from the

Judgments of Verbal Anxiety Scale revealed that hostile behavior on the part of a client-actor created more anxiety in the two groups of counselors than did friendly behavior. The levels of significance obtained were much greater than the level set in advance (.05 level) for rejection of the null hypothesis. Significance levels obtained were all at the .001 level for the more experienced counselors and better than the .005 level for the less experienced counselors. For the relatively inexperienced counselors, one individual from the group had results in the nonpredicted direction in each case. For both groups of counselors the trends for palmar sweating were in the expected direction, but were not statistically significant. Thus it could be said that the first hypothesis was supported.

The comparison between levels of counselor experience revealed that only one dependent measure significantly displayed the expected difference between groups of counselors. For the friendly interview, eyeblink rate indicated that the more experienced counselors were made less anxious by the client-actors' behavior than were the less experienced counselors. The total anxiety score from the Judgments of Verbal Anxiety Scale almost reached the .05 level of significance for this same type of interview. The results of the other dependent measures showed trends both in the predicted and non-predicted directions for the two types of interviews. It may be concluded that graduate training and experiences did not seem to affect the degree of anxiety manifested by the counselors.

Paired relationships between the measures were computed by means of the Spearman "rho" correlation. Combinations of scales were also evaluated for interrelationship by means of the Kendall coefficient of concordance. For these correlations the two groups of counselors were combined. For the interviews with the hostile client-actor, eyeblink was found to be significantly correlated with the "Client's" Scale of Counselor Anxiety and with the total anxiety score of the index of judgments of verbal anxiety. None of the other paired comparisons of measures for either type of interview produced correlations indicative of a statistically significant

relationship, nor did the group comparison of measures reveal significant intercorrelations.

### DISCUSSION

The hypothesis stating that student counselors, regardless of experience, display greater anxiety in interviews with hostile clients than in interviews with friendly clients seems to be supported, at least as far as the present counselors were concerned. Regarding the failure to find significant differences in the predicted direction between the experienced and inexperienced counselors for both friendly and hostile interviews, it is felt that the indices might not have been sensitive enough to show existing differences. The supposition that there should be differences in reactions to both types of interviews, depending on the amount of clinical experience, still seems reasonable because the idea has at least face validity, and there were more trends in the expected direction than the converse. Perhaps a more appropriate measure of experience would have been the amount of such experience which constituted work with hostile clients, specifically.

Increasing the size of the groups or extending the length of the interviews might help to distinguish between such groups from a graduate student population. Of even greater importance would be the selecting of therapist groups with a minimum of overlap between those with experience and those who were inexperienced.

Consideration of the relative value of the dependent measures as research instruments suggests that the palmar-sweating index is open to serious question as a meaningful research tool.

Eyeblink rate appeared to be quite useful for investigations of counselor anxiety. It not only discriminated very highly between types of client-actor affect for both counselor groups, but also reflected differences in direction for a particular type of interview.

The "Client's" Scale of Counselor Anxiety was also found useful, but certain modifications of the instrument appear to be necessary. The statement might be made that the differences obtained on this scale were due to biases of the actors. However, an evaluation of interactor agreement in the applica-

tion of the scale suggests that the counter-balanced design should serve to randomize any consistent bias on the part of the client-actors. Additional items need to be constructed which would allow finer discriminations to be made at the nonanxious end of the scale.

Considering the Judgments of Verbal Anxiety Scale, there were high discriminating ability of the instrument between types of interviews, and high interjudge agreement achieved in using this scale. Perhaps certain individual criteria would prove to be more useful in specific contexts; however, removing some of these criteria might cause the scale to lose its differentiating ability. Increases in numbers of counselors or the length of interviews would be possible methods of improving the discrimination of this scale even further.

As far as the lack of demonstrated relationships among the indices is concerned, it can be maintained that this did not necessarily indicate that they were too unprecise to be of value. It is suggested that instead of the four apparently independent dimensions of anxiety having separate antecedents, they might be related to a single antecedent, viz., the mediating anxiety response.

### REFERENCES

- BANDURA, A., LIPSHER, D., & MILLER, PAULA E. Psychotherapists' approach-avoidance reactions to patients' expressions of hostility. *J. consult. Psychol.*, 1960, 24, 1-8.
- BOOMER, D. S., & GOODRICH, D. W. Speech disturbance and judged anxiety. *J. consult. Psychol.*, 1961, 25, 160-164.
- CROSS, R. L., DODDS, M. E., & KNIGHTS, E. M., JR. Sudomotor test. *M. D. med. News Mag.*, 1960, 4, 65.
- DIBNER, A. S. Cue-counting: A measure of anxiety in interviews. *J. consult. Psychol.*, 1956, 20, 475-478.
- DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, 42, 3-32.
- HELLER, K., MYERS, R. A., & KLINE, LINDA V. Interviewer behavior as a function of standardized client roles. *J. consult. Psychol.*, 1963, 27, 117-122.
- KAUFFMAN, P. E., & RAIMY, V. C. Two methods of assessing therapeutic progress. *J. abnorm. soc. Psychol.*, 1949, 44, 379-385.
- KOUNIN, J., POLANSKY, N., BIDDLE, B., COBURN, H., & FENN, A. Experimental studies of clients' reactions to initial interviews. *Hum. Relat.*, 1956, 9, 265-293.

- KRAUSE, M. S. Anxiety in verbal behavior: An inter-correlational study. *J. consult. Psychol.*, 1961, 25, 272.
- LASSWELL, H. D. Verbal references and physiological changes during the psychoanalytic interview: A preliminary communication. *Psychoanal. Rev.*, 1935, 22, 10-24.
- LIGHT, B. H. Tension changes in patients undergoing psychotherapy. Unpublished doctoral dissertation, University of Illinois, 1951.
- MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, 53, 1-15.
- MAHL, G. F. Measuring the patient's anxiety during interviews from "expressive" aspects of his speech. *Trans. N. Y. Acad. Sci., Ser. 2*, 1959, 21, 249-257.
- MOWRER, O. H. *Psychotherapy theory and research*. New York: Ronald Press, 1953.
- SANFORD, F. H. Speech and personality. *Psychol. Bull.*, 1942, 39, 811-845.
- SNYDER, W. U. *The psychotherapy relationship*. New York: Macmillan, 1961.
- TIRNAUER, L. Anxiety and the behavior of psychotherapists in an experimental setting. Unpublished doctoral dissertation, Pennsylvania State University, 1959.
- ZIMMERMAN, J. Modification of the DRQ as a measure of progress in counseling. Unpublished master's thesis, University of Chicago, 1950.

(Received December 28, 1962)



## NOTES AND COMMENTS

### CHILDHOOD AND INTERCURRENT INTELLECTUAL PERFORMANCE OF ADULT SCHIZOPHRENICS<sup>1</sup>

GEORGE W. ALBEE, ELLEN A. LANE, CLARE CORCORAN, AND ANN WERNEKE

*Western Reserve University*

To determine whether intellectual deficit is a concomitant of adult schizophrenia, premorbid childhood IQs of 112 adult schizophrenics were compared with IQs obtained during the psychosis. Scores on childhood Binets and childhood group tests and on adult Wechsler-Bellevues were treated statistically to increase comparability. No significant difference was found between early premorbid IQs and scores during the psychosis. Correlations of IQs from these 2 periods of time were .55 and higher. These results seriously challenge the belief that intellectual loss occurs as a consequence of adult schizophrenia. Although hospitalized schizophrenics test below average, these same people are found to have made similar low scores on intelligence tests during childhood, long before the acute onset of schizophrenia.

Studies of intellectual deficit in schizophrenia have most often compared hospitalized schizophrenics with normal controls. In a review of these studies, Hunt and Cofer (1944) concluded that ideally "efficiency should be measured before and after the onset of the condition." Eight years later Howard Hunt (1952) wrote, "The direct approach to measuring deterioration—comparison of premorbid test results with results at the time of diagnostic study—is generally impossible."

Rappaport and Webb (1950), however, had reported that, in a sample of 10 schizophrenics, there was a mean decline of 33 IQ points between premorbid high school intelligence tests and tests administered sometime after hospitalization. Recently Lubin, Giesekeing, and Williams (1962), in a more definitive study, compared the intellectual performance of 159 schizophrenics, tested with the Army Classification Battery shortly after hospitalization, with their premorbid performance on the same test at the time of induction. They found a statistically significant, but much smaller, deficit on the second test than that reported by Rappaport and Webb. They accounted for this difference by the fact that some of the subjects in the earlier study were tested after long hospitalization.

It is generally agreed that motivational decrement is a common characteristic of schizophrenia and should be considered in studies of intel-

lectual deterioration. The use of rigidly timed tests, such as the Army Battery, that place a premium on persistence and focused attention might be expected to result in reduced scores for the schizophrenic. Most of the evidence supports Hunt and Cofer's (1944) statements that "deficit becomes greater as the complexity of the task increases" and "the findings argue that the deficit in schizophrenics is one of motivation rather than one of capacity." The use of tests that do not penalize seriously occasional lapses of attention and motivation would make this factor less important.

The present study is concerned with a comparison of premorbid intelligence test performance, not from shortly before the recognized adult disorder but from childhood, with the performance on the Wechsler-Bellevue Test administered during hospitalization for schizophrenia. The tests from childhood were either the 1916 Stanford-Binet or elementary school group intelligence tests. In the case of the former, when several Binet scores were available for one subject, the highest one was chosen as the best estimate of optimal premorbid level.

#### METHOD

Subjects were 112 adult schizophrenics under 45 years of age diagnosed at any of several state and Veterans Administration hospitals and clinics in the Cleveland area. All schizophrenics were included who had been given a Wechsler-Bellevue test during their psychosis and who had been given one or more intelligence tests in the Cleveland Public Schools during childhood.

<sup>1</sup> The research reported in this paper is supported by Grant M-5186 from the National Institutes of Health, United States Department of Health, Education, and Welfare.

Since 1937 all Cleveland school children in regular second-grade classes have been given the Kuhlmann-Anderson Test. Another intelligence test, the Cleveland Classification Test, has been given since the early 1930's to children in regular sixth-grade classes. The 1916 Stanford-Binet has been given since its inception to all children in special classes, usually every 3 or 4 years, and to a large number of other children for a variety of reasons, such as: kindergarten entrance, admission from another school system, reading and academic problems, and candidacy for enrichment programs. Most of the subjects under 32 years of age, approximately half the total sample, had been given both group and individual intelligence tests and are included in the results of both.

In comparing premorbid IQs and scores during the psychosis for those subjects who had had a childhood Stanford-Binet, Wechsler-Bellevue IQ scores were converted to an equivalent 1916 Binet IQ according to Wechsler's (1944) table. In comparing childhood group tests with adult Wechslers obtained during psychosis, childhood scores were converted to deviation IQs based on the mean and sigma of all Cleveland school children in a random year on the particular test. The mean of the deviation IQs was set at 100 and the standard deviation

at 15 in order to make these tests comparable to the Wechsler-Bellevue.

Means and standard deviations were computed for the premorbid measures and for the tests during psychosis. Product-moment correlations were computed between childhood and adult psychotic scores.

### RESULTS

As can be seen in Table 1, there was practically no difference between the mean of the highest childhood Stanford-Binet tests and the mean of the Wechsler-Bellevues administered during the psychosis for 104 subjects. The former was 87.1; the latter 87.6. The correlation between these tests, taken many years apart, was .56. This correlation compares not unfavorably with correlations of .70 to .80 that Anastasi (1958) considers usual for such measures over years.

Similarly, the mean of the Kuhlmann-Anderson deviation IQs from second grade was 92.6 while the mean hospital Wechsler-Bellevue IQ was 89.4, a decrease that was not statistically significant. The correlation between these measures was .68. The mean of the scores of the sixth-grade group intelligence test was 88.1; the mean for the same people during hospitalization was 91.5. The correlation between these two measures was .55.<sup>2</sup>

It is clear that no deficit was found between premorbid childhood intelligence tests and tests administered during the psychosis. The correlation coefficients between these two estimates of intelligence were surprisingly high, both when the premorbid individual tests and the group tests were compared with the hospital test.

### DISCUSSION

The present study fails to confirm the two previous studies in the literature which reported a decline in measured intelligence when tests administered during the psychosis were compared with premorbid measures. It also contradicts the persistent clinical contention that there is intellectual deterioration in schizophrenia. Because the highest childhood Stanford-Binet Test was used in comparison with the Wechsler-Bellevue, deterioration had more than usual opportunity to be demonstrated, if it existed. Indeed, when Stanford-Binet scores obtained between ages 11 to 14 were compared to hospital scores, an actual increase in measured intel-

<sup>2</sup> Separate analyses were made for males and females and for the subtypes of schizophrenia. No significant changes in IQ between childhood tests and adult tests were found in these analyses.

TABLE 1  
CHILDHOOD TESTS COMPARED WITH TESTS  
DURING PSYCHOSIS

	Highest childhood Binet IQ	Hospital Wechsler-Bellevue converted to equivalent Binet IQ
<i>N</i>	98	98
<i>M</i>	87.1	87.6
<i>σ</i>	11.8	15.8
<i>r</i>		.50
	Second-grade Kuhlmann- Anderson deviation IQ	Hospital Wechsler-Bellevue IQ
<i>N</i>	32	32
<i>M</i>	92.6	89.4
<i>σ</i>	15.6	14.9
<i>r</i>		.68
	Sixth-grade Cleveland Classification deviation IQ	Hospital Wechsler-Bellevue IQ
<i>N</i>	39	39
<i>M</i>	88.1	91.5
<i>σ</i>	19.9	17.9
<i>r</i>		.55



ligence during the psychosis was found (Lane, 1958).

While additional research is required to reconcile these findings with the others in the literature, it seems probable that the discrepancy is related to the nature of the tests during psychosis. The Wechsler-Bellevue Test, used in this study, is not as rigidly a speed-timed test, half the subtests not being timed at all, as the Army test on which different findings were reported. The length of time subjects had been hospitalized when tests are administered, the adequacy of the premorbid testing, the attitudes and motivation during the hospital testing—all of these are factors that should be considered in future research.

Earlier studies reported intellectual deficit when schizophrenics are compared with normals. Now it seems likely that premorbid measures on future schizophrenics would show that intellectual performance has already been depressed as a consequence of long-term personality characteristics associated with incipient schizophrenia. Schizophrenics score low in the hospital but these same people always scored low, long before the recognized onset and diagnosis of the disorder. Birren (1944) found a mean IQ of 88.7 in a group of boys who later became schizophrenic and a mean IQ of 97.7 for a normal control group of boys from the same schools. Mason (1956) found premorbid AGCT scores of 368 inductees who later became schizophrenic to be significantly below the norms made by more than a quarter of a million inductees. Since schizophrenia has been found to be much more prevalent in the lowest socioeconomic levels (Hollingshead & Redlich, 1958) these results might have been influenced by those factors related to a poor environment that are known to affect intelligence test scores.

A recent study (Reuter, 1962), however, reported a significant difference between second-grade intelligence test scores of children who later became adult schizophrenics and control children from the same elementary school classes and neighborhoods. Another study (Lane & Albee, 1963) found that children destined to be adult schizophrenics were significantly below their own siblings in intelligence tests taken during the second grade.

Intellectual deficit as a characteristic of schizophrenia then does not appear to be a concomitant of the severe or acute stages of the disorder but rather as a characteristic of the

total life picture, at least from early childhood on, of those individuals who are eventually admitted to public mental hospitals.

Although the group averages were approximately the same in these comparisons between premorbid and intercurrent tests, some individuals, of course, did go down while others stayed the same or went up in IQ after hospitalization. A study is now under way to determine whether two kinds of schizophrenics contribute to these results; that is, whether those who can be identified as "process schizophrenics" can be distinguished from those considered "reactive schizophrenics" on the basis of this long-term poor performance on intelligence tests and also on the basis of the relationship between premorbid tests and intelligence test scores during the psychosis.

## REFERENCES

- ANASTASI, ANNE. *Differential psychology*. (3rd ed.) New York: Macmillan, 1958.
- BIRREN, J. E. Psychological examinations of children who later became psychotic. *J. abnorm. soc. Psychol.*, 1944, 39, 84-95.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- HUNT, H. Testing for psychological deficit. In D. Brower & Lawrence E. Abt (Eds.), *Progress in clinical psychology*. Vol. I. New York: Grune & Stratton, 1952.
- HUNT, J. McV., & COFER, C. Psychological deficit. In J. McV. Hunt, (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944.
- LANE, ELLEN. Patterns of changes in childhood intelligence tests of psychotic adults. Unpublished doctoral dissertation, Western Reserve University, 1958.
- LANE, ELLEN A., & ALBEE, G. W. Childhood intellectual development of adult schizophrenics. *J. abnorm. soc. Psychol.*, 1963, 67, 186-189.
- LUBIN, A., GIESEKING, C. F., & WILLIAMS, H. L. Direct measurement of cognitive deficit in schizophrenia. *J. consult. Psychol.*, 1962, 26, 139-143.
- MASON, C. F. Pre-illness intelligence of mental hospital patients. *J. consult. Psychol.*, 1956, 20, 297-300.
- RAPPAPORT, S. R., & WEBB, W. B. An attempt to study intellectual deterioration by pre-morbid testing. *J. consult Psychol.*, 1950, 14, 95-98.
- REUTER, JEANETTE. Intellectual deficit in schizophrenia. Unpublished doctoral dissertation, Western Reserve University, 1962.
- WECHSLER, D. *The measurement of adult intelligence*. (3rd ed.) Baltimore: Williams & Wilkins, 1944.

(Received September 14, 1962)



# FAILURE TO VALIDATE THE *Cr* AND *Sm* SCALES OF THE MMPI<sup>1</sup>

JOHN L. VOGEL<sup>2</sup>

*Baldwin-Wallace College*

Conversion reactions and somatization have much current interest. Such terms give a dynamic twist to the more nominal linkage of psyche and soma. They speak of the patient who presents himself with somatic symptoms and complaints which are considered expressive of or dynamically related to emotional conflict. In the most dramatic case, such individuals present symptoms which are obviously incapacitating but for which no organic basis is found.

The *Cr* and *Sm* scales of the MMPI, developed by Rosen and described by Dahlstrom and Welsh (1960), were intended to differentiate individuals in whom conversion reaction or somatization appears to be operative. In an effort to validate these scales, MMPI profiles were selected from two discrete populations:

1. Psychosomatic (PS): Patients hospitalized on a medical service with somatic complaints for which no organic basis had been found at the time of observation by the psychiatric consultant. This group included 9 males and 16 females, ranging in age from 18 to 56, with a mean age of 40.6 years.

2. Somatic (S): Patients hospitalized on a medical service whose somatic complaints were consistent with findings of organic disease. This group included 13 males and 12 females, ranging

in age from 20 to 67, with a mean age of 45.7 years.

Distribution of scores, means, and standard deviations on the two scales for the two groups is shown in Table 1. Using the median as a cutting score the *Cr* scale correctly identifies 43% of the psychosomatic cases. By a similar method, the *Sm* scale correctly identifies 52%

TABLE 1  
CONVERSION REACTION AND SOMATIZATION SCALE  
SCORES IN TWO GROUPS OF PATIENTS

	<i>Cr</i>		<i>Sm</i>	
	PS	S	PS	S
65-69	6	3	31-32	1
60-64	4	10	29-30	2
55-59	5	4	27-28	5
50-54	5	4	25-26	7
45-49	2	1	23-24	3
40-44	2		21-22	5
35-39	1		19-20	1
30-34		2	17-18	1
25-29			15-16	1
20-24		1	13-14	
			11-12	1
<i>M</i>	56.3	55.40	24.88	24.00
<i>SD</i>	8.89	11.03	3.37	4.10

of the psychosomatic cases. It is apparent that the *Cr* and *Sm* scales fail to discriminate effectively in the present sample.

## REFERENCE

- DAHLSTROM, W. G., & WELSH, G. S. (Eds.) *An MMPI handbook: A guide to clinical practice and research*. Minneapolis: Univer. Minnesota Press, 1960.

(Received October 31, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from John L. Vogel (Baldwin-Wallace College; Berea, Ohio) or for a fee from the American Documentation Institute. Order Document No. 7505 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> This study was conducted while the author was at the University of Washington School of Medicine, Department of Psychiatry.

## BRIEF REPORTS

### SPONTANEOUS GSR AND ANXIETY LEVEL IN SOCIOPATHIC DELINQUENTS<sup>1</sup>

ROBERT FOX

*Longview State Hospital, Cincinnati*

AND

WALTER LIPPERT

*Hamilton County Juvenile Court, Ohio*

Recent research on autonomic function has revealed that frequency of spontaneous changes in skin resistance and in other autonomic systems is a reliable characteristic of the individual, which is relatively independent of magnitude measures (Wilson & Dykman, 1960).

The mechanism underlying emission of these responses is not completely known, nor have the behavioral correlates of differences in frequency been fully explored. However, some sets of data (Burch & Greiner, 1960) suggest the hypothesis that rate of emission is some positive function of an individual's state of "arousal" or "activation."

As a preliminary test of this proposition, this experiment measured the spontaneous activity (SA) and basal skin conductance (BSC) levels of two groups which differed in their clinically assessed level of manifest anxiety. The assumption was made that group differences in manifest anxiety level can be interpreted as indicating correlative differences in arousal intensity.

Ten male inmates of the Juvenile Court Detention Center who had received the diagnosis Sociopathic Personality Disturbance, Antisocial Reaction, formed one group (SP group), while 10 comparable subjects who had received the diagnosis Personality Pattern Disturbance, Inadequate Personality, formed a second group (IP group). Psychiatric and psychological evaluations of the manifest anxiety displayed by

these subjects indicated low anxiety levels for the SP group and high anxiety levels for the IP group. Differences between age, IQ, race, and duration of internment were not significant.

Conductance (bi-palmer) was continuously measured via a Fels Dermohmeter and Esterline-Angus recorder. Each subject, under instructions to relax, reclined on a couch located in a very low ambient noise level room. After a 15-minute adaptation period, a 10-minute recording was taken. To analyze SA, 20-second segments of each subject's 10-minute record were examined for decrements in resistance exceeding an arbitrary criterion; the total number of segments containing one or more changes constituted the score. To determine BSC level, the mean square root conductance values were taken at each minute of the 10-minute period. Results were: (a) the IP group exhibited a significantly greater frequency of SA ( $M = 14.7$ ) than the SP group ( $M = 8.5$ ),  $p = .02$ ; (b) a trend analysis of variance of the BSC data revealed no significant differences between groups and no significant trend. The correlation between SA and BSC for the SP group was  $+ .37$  ( $p > .05$ ); for the IP group  $-.62$  ( $p < .05$ ).

Absence of differences in the BSC data was not unexpected, under the conditions of the experiment. The significant difference in SA is consistent with the hypothesis that this phenomenon is positively related to some aspect of arousal.

#### REFERENCES

- BURCH, W. R., & GREINER, T. H. A bioelectric scale of human alertness: Concurrent recordings of the EEG and GSR. *Psychiat. res. Rep.*, 1960, No. 12, 183-193.
- WILSON, J. W. D., & DYKMAN, R. A. Background autonomic activity in medical students. *J. comp. physiol. Psychol.*, 1960, 53, 405-411.

(Received October 10, 1962)

## SOCIAL DESIRABILITY RESPONSE SET IN THE INDIVIDUAL<sup>1</sup>

ERLING E. BOE

*Victoria College*

AND

WILLIAM S. KOGAN<sup>2</sup>

*Veterans Administration Hospital, Seattle*

A number of studies have demonstrated high correlations ( $r$ 's in the .80s) between average Social Desirability (SD) scale values and the probability of item endorsement. However, results based on grouped data may not represent the magnitude of the SD response set in the individual subject. The results of Taylor (1959) suggest that the usual practice of grouping data serves to obscure a large amount of individual variance, although, as Taylor (1959) observed, this interpretation requires qualification in the absence of reliability data and the lengthy time interval between self-descriptions and SD ratings in his study. The problem of the present study is to investigate the magnitude of the SD response set (as measured by a point-biserial correlation) in the individual without the limitations of the Taylor study.

The distribution of SD scale values of items included in several personality inventories was found to be bimodal. Because heterogeneous samples tend to produce spurious correlation, a scale of 100 MMPI items was constructed so that their SD scale values were normally distributed about a mean representing neutral SD.

Three groups of college freshmen were formed as follows: Group I,  $N = 112$ ; Group II,  $N = 38$ ; and Group III,  $N = 36$ . Group I gave self-descriptions (yes/no) on the items; 1 week later they made SD ratings on the same items. Group II gave two self-descriptions, 1 week apart. Group III made two SD ratings, 1 week apart. Point-biserial correlations between the self-descriptions of each individual in

Group I and (a) their own SD ratings and (b) mean SD ratings yielded measures of individual SD response set. Reliabilities of individual performance, based on the subjects in Groups II and III, were used to compute a correction for attenuation.

Correlations were computed across the 100 items for each subject and averaged by using Fisher's  $z'$ . For Group I, the uncorrected mean  $r_{pb}$  between self-descriptions and group SD ratings was .27, while corrected it was .32; the uncorrected mean  $r_{pb}$  between self-descriptions and individual SD ratings was .40, while corrected it was .54; and the uncorrected mean  $r$  between group SD ratings and individual SD ratings was .49. For Group II, the mean  $r_{\theta}$  reliability coefficient was .77. For Group III, the mean  $r$  reliability coefficient was .72. The mean correlation of self-descriptions with individual SD was found to be significantly higher than the mean correlation of self-descriptions with group SD at the .001 level of confidence. Finally, the correlation between the probability of endorsement and the group SD scale values was .53.

The results suggest that correlations between probability of endorsement and group SD scale values do not accurately represent the magnitude of correlations between individual self-descriptions and group SD scale values. However, correlations between an individual's self-descriptions and his own SD ratings were significantly higher than with group SD scale values. Hence, a relevant consideration in establishing a measure of the SD response set in the individual would seem to be the basis, individual or group, for determining the SD attribute of test items. Due to the special nature of the scale used in the present study, it would be reasonable to expect the magnitude of the SD response set in the individual to be somewhat higher for some personality inventories in general use.

### REFERENCE

- TAYLOR, J. B. Social desirability and MMPI performance: The individual case. *J. consult. Psychol.*, 1959, 23, 514-517.

(Received November 1, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Erling E. Boe (Department of Psychology, Victoria College, Victoria, British Columbia, Canada) or for a fee from the American Documentation Institute. Order Document No. 7509 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> Also affiliated at the University of Washington Medical School.



## TAT AND MMPI PSYCHOPATH DEVIANT SCALE DIFFERENCES BETWEEN DELINQUENT AND NONDELINQUENT ADOLESCENTS<sup>1</sup>

ALBERT W. SILVER

*Wayne County Clinic for Child Study, Detroit*

This study sought to determine whether the performance of male adolescent delinquents selected for having characteristics associated with psychopathy differed significantly from three control groups on the MMPI Psychopath Deviant Scale (*Pd*) and the TAT.

Twenty reform school subjects were selected on the basis of a history of more than 18 months of offenses, generally unsatisfactory relationships, absence of intrapsychic tension such as anxiety, guilt, or depression, as well as of other neurotic or psychotic symptoms. Primarily environmental delinquents were excluded. This group was matched for age and intelligence with a group of mild offenders from the same institution to control for environmental factors, with a group of orphan home residents to control for the high incidence of broken homes in the psychopathic group, and with a group of high school subjects. Six TAT pictures and the *Pd* Scale were administered in small groups.

On the *Pd* Scale psychopaths had the highest number of deviant responses ( $M = 26$ ), reform school control subjects were next ( $M = 22.4$ ), followed by orphan home subjects ( $M = 22.2$ ), and by high school subjects ( $M = 18.1$ ). All possible comparisons between psychopaths and the other groups were statistically significant.

TAT protocols were analyzed by Stein and Aron's (Heymann, 1951) analysis of press and need variables. Orphan home and high school subjects told longer stories with more needs and presses than reform school subjects. Five of 33

need variables and 3 of 25 press variables discriminated significantly between at least two groups. High school subjects expressed significantly more achievement, recognition, and intro-punitive needs than both reform school groups. High school subjects also expressed more retention need and dominance press than any other group.

Reform school control subjects expressed significantly fewer rejection press and sex strivings and more affiliation press to a nonsignificant degree than any other group. On the TAT, psychopaths emerged as distinct from other reform school inmates only in expressing significantly more need sex and press rejection.

In conclusion, on an objective test (*Pd* Scale), psychopaths emerged as a distinctive group even in comparison with other reform school inmates. On the TAT, psychopaths expressed sex needs more openly but their stories were short and achievement strivings, guilt feelings, recognition, and retentive strivings were expressed less often than in the high school group. More hostile, destructive, or dependent needs did not occur in the TAT stories in psychopaths. While the high school group seemed to be the most psychologically healthy, reform school control subjects seemed to be most constricted, relying on repressive, withdrawal and avoidant defenses for maintaining some degree of social conformity. The concept of psychopathy seems to have some validity even when applied to adolescents with relatively brief delinquent histories. The limited fantasy resources of psychopaths, however, argue against the use of an instrument which calls for fantasy expression in understanding their psychodynamics.

### REFERENCE

- HEYMANN, G. M. A TAT investigation of some psychological factors in Acne. Unpublished master's thesis, University of Wisconsin, 1951.

(Received November 8, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Albert Silver, Wayne County Clinic for Child Study, 1025 East Forest Avenue, Detroit 7, Michigan, or for a fee from the American Documentation Institute. Order Document No. 7510 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$2.50 for microfilm or \$1.75 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

## DEVELOPMENT OF POTENTIALLY THERAPEUTIC INTERPERSONAL RELATIONS IN SCHIZOPHRENICS THROUGH COMPETITION<sup>1</sup>

ALBERT E. MYERS<sup>2</sup>

*Educational Testing Service, Princeton, New Jersey*

This study investigated the effects of team competition on the social behavior of schizophrenic patients. Earlier, Myers (1962) found that interteam competition led to better adjustment of team members than did an otherwise comparable noncompetitive condition. This finding suggested the possibility of using team competition as a therapeutic device in the treatment of people with mental disorders.

Forty-eight chronic schizophrenic male patients were selected to participate in a 6-week project. The patients were easily grouped into three levels of psychological incapacitation.

The subjects were combined into 24 two-man teams. Each team participated 1 day every week in a recreational golf program.

The first 4 weeks were devoted to practice. Two teams were combined to form a foursome and played an average of six holes per day in a noncompetitive fashion. The teams were then randomly assigned to competitive and noncompetitive conditions for the final 2 weeks. Verbal interactions of the patients were recorded while they played.

Establishing a criterion in this kind of research is, by no means, a simple task. The typical "Hospital Adjustment Scale" was useless; interview criterion data proved fruitless;

paper-and-pencil measures were out of the question.

Consequently, a whole new approach to criterion assessment was called for. The method which was chosen was a unique one and, hence, represented somewhat of an experimental gamble. It seems that this approach perhaps deserves serious consideration by researchers in this area.

The basic assumption was made that patients with a relatively less severe degree of incapacitation enjoyed better interpersonal relations. That is, there should be a strong association between the level of pathology of a patient and his ability to function adequately in a group. A positive change in interpersonal adequacy or adjustment is, then, operationally defined as a team behaving more similarly to the better adjusted patients at the end of the experience than it did at the beginning.

With these assumptions, multiple correlations were run with adjustment ratings as the criterion. This provided weights for the best predictors so as to permit the compilation of a single score. This score represented the particular weighting of those behaviors which were maximally effective in discriminating across the range of pathology. A spontaneous change in the appropriate behaviors was, once again, assumed to index a change in the level of interpersonal adequacy. The results suggested that competition facilitated adjustment among the better patients but that it had an unfavorable effect among the poorer patients.

Our apparent success with verbal behavior suggests that certain inhibitions toward using this type of data with schizophrenics is unwarranted.

### REFERENCE

- MYERS, A. E. Team competition, success, and the adjustment of group members. *J. abnorm. soc. Psychol.*, 1962, 65, 325-332.

(Received November 9, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Albert E. Myers, Educational Testing Service, Princeton, New Jersey, or for a fee from the American Documentation Institute. Order Document No. 7511 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> Carried out with the assistance of the Group Effectiveness Research Laboratory (Fred E. Fiedler, Principal Investigator) under contract DA-49-007-MD-2060 of the Office of Surgeon General, Department of the Army, and the Veterans Administration Hospital, Danville, Illinois.

KENT E-G-Y:  
DIFFERENTIAL SCORING AND CORRELATION WITH THE WAIS<sup>1</sup>

GERALD L. CLORE, JR.<sup>2</sup>

*University of Texas*

The Kent E-G-Y Test has been widely used to estimate the level of intellectual functioning of hospital patients. The test consists of only 10 items, and the author of the test gives the individual examiner a wide latitude in the interpretation and scoring of the items. The present study was carried out to determine the effect of different degrees of strictness in scoring the Kent on final Kent scores and their relation to WAIS scores.

From a general medical and surgical Veterans Hospital population a representative sample of 30 psychological referrals was tested. Two scorings of the Kent were made. An easy score was obtained by interpreting questions liberally and by crediting ambiguous answers, highest order abstractions, and correct answers that were spoiled by incorrect additional answers. A strict score resulted from interpreting questions strictly and from crediting only clearly correct answers, lowest order abstractions, and only unspoiled correct answers.

<sup>1</sup> An extended report of this study may be obtained without charge from Gerald L. Clore, Jr. (Department of Psychology, University of Texas, Austin 12, Texas) or for a fee from the American Documentation Institute. Order Document No. 7512 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> This study was conducted at the Veterans Administration Hospital at Dallas, Texas. The author is grateful for helpful suggestions from A. J. Jernigan, Chief Psychologist.

The mean score of the Kents which were scored easily was 26.6 and the mean score of the strictly scored Kents was 24.1. A *t* test for difference between correlated means yielded a *t* value of 6.63 which was significant at the .001 level with 29 *df*. Pearson product-moment correlations between easy Kent scores and WAIS scores were: Full Scale .68, Verbal .69, Performance .61. Correlations between strict Kent scores and the WAIS were: Full Scale .70, Verbal .69, Performance .63. All six coefficients were statistically significant at the .01 level. To determine if the three easy Kent correlations differed from the three strict Kent correlations, a test for differences between correlations on the same sample was used. There were no significant differences between the correlations.

Since the mean Kent scores were significantly different, but the correlations of easily and strictly scored Kents with WAIS scales were not significantly different, it is important to score the Kent using the same procedure that was used when the norms for the particular population tested were obtained. However, different scoring procedures do not affect the Kent's validity. The size of the correlations indicate that the Kent is a useful estimator of WAIS scores. It should be noted that the Kent has no correction for age, while WAIS IQs are corrected for age differences. This factor would, however, be expected to work against a high correlation, so an even higher correlation might be expected if WAIS weighted scores were used rather than IQs.

(Received November 13, 1962)



## EXPERIMENTAL PREDICTION OF THE LONG-TERM EFFECT OF PSYCHIATRIC THERAPY<sup>1</sup>

GUDMUND J. W. SMITH AND GUNNAR JOHNSON

*Psychological Laboratory, Lund, Sweden*

In previous investigations (see Smith & Johnson, 1962) the present authors have described how a serial (microgenetic) tachistoscopic technique utilizing pairs of stimuli (one being incongruent with or threatening to the other) was exploited for the study of the effects of psychiatric therapy. Reactions to this stimulation were scored in three main categories of pre-cognitive (defensive) organization: *isolation-negation*, *repression*, and *projection*. A fourth main category referred to a general syndrome of *depression* as revealed in the perceptual genesis, a fifth to signs of discontinuity signifying disturbances of a *psychotic* order. Parallel versions were used before and after therapy. The results implied that isolation and repression were hardly affected by therapy whereas many patients with depression or slight projective tendencies seemed to recover. Psychotic patients were often registered as improved but seldom as recovered. The present authors also wanted to explore, 2 years later, whether the original scoring categories could predict the subsequent destinies of the patients.

The original 48 patients could be divided into two groups. In one group (U) either no improvement was reported for the subjects, or they relapsed within half a year. The other group (Im) consisted of patients described as improved and not reporting any serious complaints within the first 6 months, most of them not within 2 years. This differentiation resulted in two groups of 24 patients each. We only distinguish here between presence of a sign, and absence. Since isolation and repression had proved to be quite stable, they were scored even if they appeared only in one of the experiments. In the case of the more unstable signs of projection and depression, results obtained before therapy were used for making predictions.

Among 19 cases with isolation 13 belonged to Group Im. The result certainly does not mean

that their compulsive symptoms disappeared or even improved substantially. But the compulsive character is probably likely to accept his status after therapy as a conclusive sign of how much he is able to improve. The effects of ego support in a hysteroid person, on the contrary, are often known to be no more lasting than traces in sand. An exception was made for patients with simultaneous signs of isolation-negation. Thus, 8 of 9 patients with repression but no isolation belonged to Group U, but only 1 of 6 patients with isolation plus repression. In the case of projection we picked out one category previously marked as grave and predicted that the probability of a stable improvement would be lower than in other subjects with projection. None of the mild cases showed simultaneous signs of repression. Group Im, then, included 11 of 16 mild cases but only 1 of 8 severe ones. If signs of depression before therapy appeared together with signs of psychotic discontinuity, the probability of a cyclic recurrence of the symptoms would be quite high. One category of 11 patients with depressive signs plus signs of discontinuity (6 patients) or repression (5 patients) but no signs of isolation was compared with another category of the 16 remaining cases. Among 13 individuals in Im, 12 belonged to the latter category.

An overall prediction for Group U, those who relapse easily or do not even improve in the clinic, would utilize the presence of severe signs of depression and projection, signs of repression but absence of signs of isolation. The rest of the patients would be more likely to belong to Group Im. There were 17 patients with negative signs in Group U and 2 in Group Im ( $\chi^2 = 19.60$ ,  $p < .001$ ). If severe signs of projection or depression are allowed to dominate simultaneous signs of isolation the numbers are but little affected.

### REFERENCE

- SMITH, G. J. W., & JOHNSON, G. The influence of psychiatric treatment upon the process of reality construction: An investigation utilizing the results of a serial tachistoscopic experiment. *J. consult. Psychol.*, 1962, 26, 520-526.

(Received November 28, 1962)

<sup>1</sup> An extended report of this study has been deposited with the American Documentation Institute. Order Document No. 7513 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

## ERRATA

In the article by Edmund S. Howe and Benjamin Pope, "Therapist Verbal Activity Level and Diagnostic Utility of Patient Verbal Responses" (*J. consult. Psychol.*, 1962, 26, 149-155), two corrections are required in the right column of page 153. In line 5 the word "positive" should read "negative"; in line 6 the word "negatively" should read "positively."

In the article by Irving Steingart, "Conditions of Personality Organization Related to Empathic Ability Among Normal and Schizophrenic Adults" (*J. consult. Psychol.*, 1962, 26, 416-421), the second half of the sentence on page 418 beginning with the words "The schizophrenic 'high empathizer' group" should read: "the schizophrenic 'low empathy' group consisted of eight subjects who obtained empathy scores of 0-4 (the lower 17.4%)."

In the article by John E. Exner, Jr., Eugene McDowell, Joan Pabst, William Stackman, and Lynn Kirk, "On the Detection of Willful Falsification in the MMPI" (*J. consult. Psychol.*, 1963, 27, 91-94), the statement, "Gynther (1961) has demonstrated that a high F score is found in almost all cases of behavioral disorders . . .," should have read as follows: "Gynther (1961) found that in a group of white males referred by the court for psychiatric examination, almost all the high F scores obtained were given by cases of behavioral disorders."

# PSYCHIATRY

JOURNAL FOR THE STUDY OF INTERPERSONAL  
PROCESSES

CONTENTS OF NO. 1, VOL. 26

FEBRUARY, 1963

THEODORE LIDZ AND OTHERS: Schizophrenic Patients and Their Siblings.

PHILIP M. MARGOLIS AND OTHERS: The Patient-Staff Meeting—A Technique for Encouraging Communication in the Psychiatric Hospital.

JOSEPH D. LICHTENBERG: The Return to Reality as a Critical Phase in the Treatment of Schizophrenic Patients.

HENRY U. GRUNEBAUM AND OTHERS: The Baby on the Ward.

STEWART E. PERRY: Home Treatment and the Social System of Psychiatry.

ROBERT N. BUTLER: The Life Review: An Interpretation of Reminiscence in the Aged.

ROGER L. SHAPIRO: Adolescence and the Psychology of the Ego.

IRVING SCHNEIDER: The Use of Patients to Act Out Professional Conflicts.

JOHN T. MONROE, JR. and D. WILFRED ABSE: The Psychopathology of Trichotillomania and Trichophagy.

Book Reviews

\* \* \*

Back issues are available for \$2.50 (\$2.70 foreign) for the years 1954-1962, and for \$2.00 (\$2.20 foreign) for the years 1938-1953.

A discount of 20 percent is given on orders of twenty or more back issues.

William Alanson White Psychiatric Foundation  
1610 New Hampshire Avenue, N.W., Washington 9, D.C.

*Subscriptions \$10.00, foreign \$10.80*



# Volume II

**Hans H. Strupp  
Lester Luborsky**  
*Editors*

of

# RESEARCH in PSYCHOTHERAPY

Proceedings of the Second Conference on Research in Psychotherapy held at Chapel Hill, N.C., in May 1961. Leading research workers in psychology and psychiatry deal with selected research issues and assess new research developments. Their efforts place research in psychotherapy more squarely within psychological research concerned with understanding interpersonal processes.

**The psychotherapist's  
contribution to the treatment process**

**Measuring personality change  
in psychotherapy**

**Definition, measurement, and  
analysis of significant  
variables**

The book is a follow up of *Research in Psychotherapy*, published by the American Psychological Association in 1959, which reported on the First Conference on Research in Psychotherapy held in Washington, D.C., in 1958.

## Contributors

Barbara J. Betz  
Jack Block  
Edward S. Bordin  
John M. Butler  
Kenneth Mark Colby  
Jerome D. Frank  
Ernest A. Haggard  
Leonard Krasner  
Henry L. Lennard  
Daniel J. Levinson  
Maurice Lorr  
Lester Luborsky  
Morris B. Parloff  
Laura N. Rice  
Irwin C. Rosen  
Nevitt Sanford  
George Saslow  
David Shakow  
John M. Shlien  
Richard S. Siegal  
William U. Snyder  
Ian Stevenson  
Hans H. Strupp  
Alice K. Wagstaff

329 pp.  
plus Indexes  
\$3.50

**Order from:**

**AMERICAN PSYCHOLOGICAL ASSOCIATION**  
1333-16th St., N.W., Washington 6, D.C.

# JOURNAL OF CONSULTING PSYCHOLOGY

October 1963

Vol. 27, No. 5

## Contents

### ARTICLES

A Study of Certain Schizophrenic Dimensions and Their Relationship to Double Alternation Learning: <i>Walter J. Johannsen, Samuel H. Friedman, Thomas H. Leitschuh, and Helen Ammons</i>	375
The Differential Performance of Orals and Anals in a Verbal Conditioning Paradigm: <i>Edwin O. Timmons and Charles D. Noblin</i>	383
Frequency Hypothesis and Content Analysis of Projective Techniques: <i>John Wallace and Lee Sechrest</i>	387
The Relationship of Expectancy of Reward to Achievement Performance on an Arithmetic and Thematic Test: <i>Bernard I. Murstein</i>	394
Extraexperimental Effects in Verbal Conditioning: <i>William R. Wimsatt and Norris D. Vestre</i>	400
Counselor Attitudes and Client Behavior: <i>Irene E. Waskow</i>	405
The Construct Validity of the Progressive Matrices as a Measure of Superego Strength in Juvenile Delinquents: <i>Wilfred H. Higashimachi</i>	413
Factors of Psychopathology in the Ward Behavior of Acute Schizophrenics: <i>Allen Raskin and Dean H. Clyde</i>	420
Interpersonal Orientation in Relation to Hypnotic Susceptibility: <i>P. M. Bentler</i>	426
Response Set Paralysis: Implications for Measurement and Control: <i>Jerome Siller and Abram Chipman</i>	432
Comparative Analysis of Psychotic Depressives with Matched Normals on Some Untimed Verbal Intelligence Tests: <i>Samuel Granick</i>	439
Erratum	443

### NOTES AND COMMENTS

Use of the Shipley-Hartford Conceptual Quotient as a Measure of Intellectual Impairment: <i>Peter M. Lewinsohn</i>	444
Overinclusive Thinking in Process and Reactive Schizophrenics: <i>Thomas S. Eliseo</i>	447
Ecology of Service: <i>Haim G. Ginott and Dell Lebo</i>	450
Method for the Study of Staff Interaction through Ratings of Patients: <i>Jerome Fisher</i>	453
Use of the Terman-Merrill Abbreviated Scale on the 1960 Stanford-Binet Form L-M on Negro Elementary School Children of the Southeastern United States: <i>Wallace A. Kennedy, Vernon Van De Riet, and James C. White, Jr.</i>	456

### BRIEF REPORTS

Social Desirability: A Reply to Wiggins: <i>Jerald N. Walker</i>	458
Reliability of Measurement under Various Experimental Instructions: <i>Margery H. Krieger and Charles B. Schultz</i>	459
The Relation of the Gordon Personal Inventory to Several External Criteria: <i>Warren W. Willingham and Rosalie K. Ambler</i>	460
Personality Variables in the DAP: <i>Marjorie Heberlein and F. I. Marcuse</i>	461
Differentiating Psychotics from Organics on the Minnesota Percepto-Diagnostic Test: <i>Ensley Uyeno</i>	462
The Relationship between Depressive Personality Characteristics and Rorschach Card Preference: A Reply to T. E. Smith: <i>Stephen Neiger and Christos Papastergiou</i>	463
Need for Social Approval as Reflected on the TAT: <i>Russell P. Norman</i>	464
Dimensions of Personality Adjustment in Mental Patients: <i>John G. Hurst and Gordon Olson</i>	465
A Comparison between Taylor's and Freeman's Manifest Anxiety Scales: <i>Iris Balshan Goldstein</i>	466

# Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

## Specifications

**Brief Report.** The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the

author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

<sup>1</sup> An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Documentation No. — from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$— for microfilm or \$— for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

**Extended report.** Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (a) The extended report should be typed with single spacing for economy in duplication. (b) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (c) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.



## A STUDY OF CERTAIN SCHIZOPHRENIC DIMENSIONS AND THEIR RELATIONSHIP TO DOUBLE ALTERNATION LEARNING

WALTER J. JOHANNSEN, SAMUEL H. FRIEDMAN,  
THOMAS H. LEITSCHUH, AND HELEN AMMONS<sup>1</sup>

*Veterans Administration Center, Wood, Wisconsin<sup>2</sup>  
and Marquette University School of Medicine*

In order to clarify relationships among four dimensions thought to underlie schizophrenia, 52 patients were rated on a scale of premorbid personality and on a scale measuring the process-reactive dimension. They were further categorized with respect to chronicity and the presence of paranoia. Results indicated that the process-reactive, good-poor premorbid personality, and acute-chronic dimensions are essentially similar; and that the paranoid-nonparanoid dimension is independent of the others. Testing on a double alternation learning task demonstrated differences between paranoids and nonparanoids, but not between the poles of the other dimensions.

In recent years, several researchers have advocated a change in the present nosology of schizophrenia. Suggestions have included symptom-based systems derived from factor analysis (Guertin, 1952; Lorr, McNair, Klett, & Lasky, 1962; Wittenborn, 1951), but none of these methods has found wide use. There is another empirical approach, in which assumptions have never been made explicit, but which has nonetheless generated considerable research. This method consists of postulating classificatory dimensions along which all schizophrenics can be placed and then relating the scores on these dimensions to behavioral correlates. The utility of the posited dimension then depends on the range and importance of behavior correlated with it. Once these dimensions are firmly grounded in stable observable events, etiological and prognostic correlates can be elicited.

Research has already demonstrated that two dimensions derived from the older nomenclature have genuine descriptive value. One of these is the paranoid-nonparanoid schizophrenic dimension. Raters using inter-

view data can reliably differentiate these two groups even when other categorizations are difficult. Paranoids have been differentiated from nonparanoids on a variety of psychological functions and abilities: tapping speed (Shakow & Huston, 1936); steadiness (Shakow & Huston, 1946); perception of Lissajou figures (Philip, 1953); Rorschach genetic level (Siegel, 1953); degree of suggestibility (Williams, 1932); conditioning (O'Connor & Rawnsley, 1959); response to incentives (Topping & O'Connor, 1960); alternation learning under social feedback conditions (Johannsen, 1961); degree of intelligence test decrement (Mason, 1956); and perception of self (Havener & Izard, 1962). Goldstein and Carr (1956) suggest that differences in parental child rearing attitudes differentiate the groups as well.

A similar picture emerges when schizophrenics are divided into acute (in the sense of recent onset) and chronic groups. Stable differences have been found between them on: reaction time (Huston & Senf, 1952); gestalt perception (Cohen, Senf, & Huston, 1956); response to praise and censure (Long, 1961); intelligence test performance (Feinberg & Garman, 1961); vocabulary (Rabin, King, & Ehrman, 1955); and appreciation of humor (Senf, Huston, & Cohen, 1956).

During the last decade considerable quantitative work has developed about two other

<sup>1</sup> Johannsen is now at the Curative Workshop, Milwaukee, Wisconsin; Friedman is at Milwaukee, Wisconsin; and Helen Ammons is at the Veterans Administration Hospital, Palo Alto, California.

<sup>2</sup> Gratitude is due to Douglas Holmes, who assisted in the running of the subjects, and to Shirley Brehm, who performed the statistical analysis of the data.

dimensions: process-reactive and good-poor premorbid personality. The history of the former category extends back to Bleuler (1950), but research in the area was given impetus by the study of Kantor, Wallner, and Winder (1953) which systematized the elements that defined the dimension. In essence, reactive schizophrenia differs from process schizophrenia in being of more sudden onset and involving more adequate interpersonal relationships, normality of affect, presence of precipitating cause and a better response to treatment. The literature reveals that process patients could be distinguished from reactive patients on: physiological response to meclolyl (King, 1958); flicker fusion threshold (McDonough, 1960); response to incentives (Reisman, 1960); Rorschach genetic level (Fine & Zimet, 1959); Rorschach organic signs (Brackbill & Fine, 1956); and performance on Benjamin Proverbs (Becker, 1956).

A fourth posited dimension, good-poor premorbid personality, is defined in terms of a scale developed by Phillips (1953), which deals with the sex life and interpersonal relationships of the patient. Due primarily to the interest engendered in this scale by researchers at Duke University, an extensive body of literature has developed. Differences between the two groups have been established in many areas: size estimation (Harris, 1957); perception of pictures with interpersonal themes (Zahn, 1959); retention of hostile themes (Mallet, 1956); and reminiscence (Bleke, 1955). Regarding etiology, differences have been found in parental child rearing attitudes (Garmezy, Clarke, & Stockman, 1961); patterns of parental dominance (Baxter & Becker, 1962); and rural versus urban residence (Query, 1961). Finally, Farina and Webb (1956) have demonstrated prognostic implications for Phillips Scale scores.

Examining the range of behavior touched upon, one notices that differences occur on similar tasks for different categorizations. Perceptual tests differentiate acutes and chronics, process and reactive, paranoids and nonparanoids, good and poor premorbid. Are these dimensions independent or are we

referring to the same underlying dimension in every case? Is a paranoid schizophrenic one who has a good premorbid personality? Is an acute schizophrenic always reactive? Becker (1959), referring to earlier research, states, "These studies have led to a number of terms, all describing the same distinction: malignant-benign, chronic-acute, process-reactive [p. 442]." Moreover, recent workers (Chapman, Day, & Burstein, 1961; DeWolfe, 1962) refer to the Phillips Premorbid Scale as a measure of process-reactive schizophrenia.

However, the issue is not so clear-cut. The item content of the Phillips Scale hardly overlaps the Kantor, Wallner, and Winder (1953) alternatives. Those who claim that acute schizophrenia is equivalent to reactive schizophrenia ignore the facts that the acute-chronic dimension refers to a temporal sequence, while the process-reactive dimension is structural, and that many long-term patients begin their schizophrenic careers with a reactive-like break with reality. Moreover, one might predict that a paranoid schizophrenic (usually considered "more intact") would be more likely to be classified as reactive; yet one of the scale alternatives defining process schizophrenia is the presence of massive paranoia.

The major purpose of this study is to examine the relationships among these four dimensions and to determine whether subjects classified as members of one category within a dimension (e.g., reactive) will also be classified as members of an analogous category within another dimension (e.g., good premorbid personality). Its secondary purpose is to extend the construct validity of these schemes by relating these dimensions to a learning task which is believed to have relevance to an important aspect of behavior, namely, the ability to utilize information about performance derived from an external human source.

## METHOD

### *Subjects*

The sample consisted of 52 consecutive schizophrenic admissions to the psychotic ward of the Wood Veterans Administration Center psychiatric service. Initial screening by the ward physician

TABLE 1  
ACTUARIAL CHARACTERISTICS OF THE SAMPLE

Group	Total hospital (mos.)	Time since onset (yrs.)	Education (yrs.)	IQ	Vocabulary score	Age	N
Total:							
<i>M</i>	21.33	7.79	11.20	105.65	28.28(9)	36.96	52
<i>SD</i>	32.77	6.24	2.86	18.76	6.31	7.17	
Acute:							
<i>M</i>	2.30	2.70	12.22	110.00(2)	27.87(2)	35.40	10
<i>SD</i>	2.05	5.10	2.39	14.80	5.13	6.07	
Chronic:							
<i>M</i>	25.74	9.00	10.98	104.66(7)	28.37(7)	37.61	42
<i>SD</i>	34.97	5.86	2.91	19.41	6.55	7.28	
Paranoid:							
<i>M</i>	23.14	9.00	11.32	113.13(6)	30.53(6)	38.75	21
<i>SD</i>	30.20	6.82	2.96	18.84	6.54	6.44	
Nonparanoid:							
<i>M</i>	19.93	6.97	11.13	101.64(3)	27.07(3)	36.16	31
<i>SD</i>	34.33	5.66	2.80	17.44	5.84	7.34	
Process:							
<i>M</i>	32.08	10.65	10.80	106.22(3)	29.56(3)	38.88	26
<i>SD</i>	40.58	6.16	3.18	19.29	6.76	7.38	
Reactive:							
<i>M</i>	10.38	4.92	11.60	105.00(6)	26.80(6)	35.54	26
<i>SD</i>	16.29	4.84	2.50	18.11	5.39	6.44	
Good:							
<i>M</i>	21.23	6.18	10.86	100.76(5)	25.53(5)	36.90	22
<i>SD</i>	43.83	5.95	2.82	16.97	4.78	6.97	
Poor:							
<i>M</i>	21.23	8.97	11.45	108.85(4)	30.08(4)	37.37	30
<i>SD</i>	21.27	6.18	2.87	19.18	6.54	7.21	

Note.—Numerals in parentheses above refer to the number of subjects missing from that cell. A total of nine subjects were discharged before completion of intelligence testing, but after ratings were obtained; three did not perform the double alternation task. Actuarial characteristics of the reduced samples were virtually identical to the larger sample.

eliminated individuals unsuitable for the study. All subjects were between the ages of 20 and 55, were clear of complicating neurological conditions, and were withheld from somatic therapy until the completion of testing. The pertinent actuarial characteristics for the total sample are presented in Table 1. Evaluation of the clinical data indicated that this population is characterized by frequent relapses, multiple hospitalizations, and comparatively brief periods of inpatient care. From the viewpoint of total length of hospitalization, the overall sample

would be comprised of relatively early schizophrenics, as compared to the bulk of Veterans Administration psychotics, although they would appear moderately chronic from the viewpoint of time since the initial diagnosis of schizophrenia.

#### Procedure

The paranoid-nonparanoid and acute-chronic dimensions were evaluated from psychiatric material, the process-reactive and good-poor premorbid personality dimensions from interview ratings.



*Paranoid-nonparanoid.* Diagnosis established by the ward staff on admission was accepted as the criterion of the paranoid-nonparanoid dimension. Since it appears that this discrimination can be made reliably (e.g., Orgel, 1957), a more refined method of categorization was deemed unnecessary. Three of the four raters who participated in the study were unaware of the subject's diagnosis at the time of rating. The fourth, a psychiatrist, participated in the admission staff meeting where diagnosis was established. Any bias resulting from the knowledge of the diagnosis was deemed minimal, since agreement on the Phillips Scale ratings between him and his fellow rater (who was unaware of the subject's diagnosis) was excellent.

*Acute-chronic.* Subjects were assigned to the acute group if total hospitalization was less than 6 months and the earliest schizophrenic diagnosis occurred within the previous year. One subject had been diagnosed schizophrenic 18 years before, but had since made a good extramural adjustment, and had remained out of the hospital during that period. Since he bore a current acute schizophrenic diagnosis, he was included in the acute group for purposes of this study. All other subjects, a total of 42, were considered chronic.

*Process-reactive.* The 24 items of the Kantor, Wallner, and Winder (1953) alternatives were arranged in two columns, one referring to the process, the other to the reactive dimension. The rater checked the alternatives that best described the subject. After all items were checked, he then rated the subject on a 5-point bipolar scale which ranged from "classic process picture" to "classic reactive picture," with a midpoint classification for individuals who showed an equal mixture of process and reactive characteristics. The rating was to be a clinical judgment, in which the clinician was not necessarily bound by the number of characteristics checked in one or the other column. This expedient was deemed preferable to an arithmetic total because it could not be assumed that alternatives were of equal discriminating power. Where disagreements between raters existed, discrepancies were resolved by totaling the number of process attributes checked by the raters, comparing this total with the number of reactive attributes checked, and assigning the subject to the category in which the majority of checks appeared. Justification for this procedure lies in the fact that in only 7 of 104 protocols was there a discrepancy between the column with the majority of check marks and the final classification assigned.

*Good-poor premorbid personality.* Section 1 of the Phillips Scale was used as a measure of premorbid personality, following the usage of the Duke group (Rodnick & Garnezy, 1957). Several areas of premorbid adjustment were scored on separate 7-point scales: recent social adjustment, social aspects of sexual adjustment during adolescence and immediately beyond, sexual adjustment above and below age 30, and history of personal relations.

Median scores were obtained, and the subject was placed in the "poor premorbid" category if his median rating fell above the scale midpoint, and in the "good premorbid" category if below.

Three psychologists and one psychiatrist served as raters. Two psychologists rated subjects on the process-reactive scale; the other psychologist and the psychiatrist rated them on the Phillips Scale. The raters were all experienced, having a minimum of 5 years work with patients in mental hospital settings. Ratings were done independently, during and after a brief, semi-structured interview, conducted by one of the psychologists and observed by the other raters. Interview ratings were deemed preferable to the more common method of rating clinical history data, since the majority of patients' folders were incomplete. To avoid the possibility of raters basing their judgments merely on a "halo effect" derived from the interviewer's techniques, the questions asked were brief, deliberately confined to the content of the two scales, and designed to elicit categorical answers (e.g., "Have any of your brothers or sisters ever been a patient in a hospital like this? Do you get into fights? Often? Recently? How old were you when you had your first sexual experience?").

Each subject was again seen, usually during the week following the interview, and administered the Shipley Hartford Intelligence Scale to insure that groups were adequately equated on the intelligence variable. Soon afterwards, they were tested on a double alternation task, described in full elsewhere (Johannsen, 1959). In brief, subjects were seated at a desk with an unpainted 6 × 10 inch wooden panel placed flat on its surface. Two doorbell buttons were mounted on this response board, 2 inches to the right and to the left of center. The experimenter sat beside the desk at a 90° angle to the subject. The response buttons were wired in a circuit to two 7.5-watt light bulbs mounted on a panel placed inconspicuously on the floor near the experimenter. These lights served to signal to the experimenter which of the two buttons had been pressed. By focusing on the lights, the experimenter was prevented from seeing the motion of the subject's hands and inadvertently giving premature information to him. All of the subject's responses were recorded.

The task was the learning of a repetitive double alternation pattern (RRLRLRL). Two hundred trials were given, each of which required the subject to respond by pressing one of the two buttons. After each of the subject's responses, the experimenter replied "right" or "wrong" in a level tone of voice. Adequate performance on this task requires an ability to codify, assimilate, and use information from an external human source. A 40-trial initial series was run without feedback in order to control for response biases and to insure ability to comprehend instructions. No subject had to be disqualified for confusion.

## RESULTS

*Relationship Among Dimensions*

To determine how reliably raters could classify subjects on the process-reactive and good-poor premorbid personality scales, the two raters on each form were compared for percentage agreement. On the Phillips Scale the psychologist and psychiatrist raters agreed on the dichotomous classification (good versus poor) in every instance. A threefold classification was used to evaluate reliability in the process-reactive scale (process, reactive, and intermediate). Here the degree of agreement between the independent ratings was 88%, sufficient to justify further analysis of the data. In data analysis the intermediate category was eliminated. All subjects falling into this group were assigned to either process or reactive categories following the procedure mentioned previously.

To evaluate the degree of relationship among the four dimensions studied,  $2 \times 2$  tables were constructed for each of the six possible combinations of dimensions. Since it is not known whether these dimensions represent normally distributed continua (e.g., that there is a gradation of "processness" from a classic process picture to a reactive picture devoid of process elements), both the tetrachoric  $r$ , which presupposes such a distribution, and the phi coefficient, which does not, were computed in each instance. The results are summarized in Table 2.

It is evident that certain significant relationships do exist between the acute-chronic and process-reactive dimensions, the acute-chronic and good-poor dimensions, and the good-poor and process-reactive dimensions. These relationships are affirmed by both correlational methods employed. It thus appears that an acute schizophrenic will be classified as reactive and as having a good premorbid history, that a chronic schizophrenic will be classified as process and as having a poor premorbid personality, that good premorbid are usually classified as reactive and poor premorbid as process. The paranoid-nonparanoid dimension appears to be independent of the other three. The high degree of relationship between the good-poor and

TABLE 2

CORRELATIONS BETWEEN VARIOUS HYPOTHETICAL DIMENSIONS OF SCHIZOPHRENIA

Comparison	$r_t$	$\phi$	$\phi \text{ max}^a$
Acute-chronic versus paranoid-nonparanoid	.18	.09	.59
Process-reactive versus paranoid-nonparanoid	.06	.04	.82
Good-poor versus paranoid-nonparanoid	.15	.09	.96
Acute-chronic versus process-reactive	.71*	.39*	.49
Acute-chronic versus good-poor	.72*	.37*	.57
Good-poor versus process-reactive	.62*	.39*	.86

\*  $p < .01$ .<sup>a</sup> Values refer to maximum phi possible with the cell frequencies involved.

process-reactive dimension can be accepted with confidence. However, the small number of acute subjects renders any conclusion involving them tentative, and must be accepted with caution.

An attempt was made by item analysis to determine which items on the process-reactive scale served as cues for raters' judgments. Chi squares were computed for each item relating them to the final categorization. Eight of the 24 items discriminated at better than the .05 level of confidence for both raters. Two additional items discriminated for one rater though not for the other. It would appear that the same set of cues was used in making the final clinical judgment, and it is probable that the scale in the form used in this study can be refined and abbreviated.<sup>3</sup>

*Relationship of Dimensions to Alternation Learning*

Prior to analyzing the differential performance of the various groups on the double alternation task, it was necessary to determine to what extent extraneous variables played a part in the results obtained. Examination of Table 1 indicated that groups were closely equated on most of the variables likely to affect performance. Where two groups showed

<sup>3</sup> Copies of the abbreviated form may be obtained from the senior author.



more than marginal differences, e.g., paranoids versus nonparanoids in intelligence, analysis of variance was used to assess the significance of these differences. Only length of hospitalization and time of onset showed significant differences between diagnostic groups. The effect of these variables was then evaluated by correlating them with scores obtained on the double alternation task. No correlation achieved significance. Drug effects were not studied, but evidence obtained earlier (Johannsen, 1961) indicated that medication has little effect on this task. No examiner differences appeared, and the subjects' preliminary response sets were unrelated to subsequent task performance. We may therefore conclude that the effect of these variables on any obtained group differences is minimal.

The basic analysis was performed on the number of correct responses made during the 200 trials administered. Four analyses of variance were performed comparing acutes versus chronics, paranoids versus nonparanoids, process versus reactive and good versus poor premorbid personality. Only the *F* test for the paranoid-nonparanoid classification was significant ( $F = 5.18, 1/47 \text{ df}, p < .05$ ), with paranoids performing more adequately than nonparanoids. The mean number of correct responses was 153.00 for paranoids, 126.07 for nonparanoids. This result is in accord with that obtained by Johannsen (1961). In the earlier study, Johannsen noted a strong relationship between perseverations and number of errors, where perseverations were defined as single or RL alternations. Here, however, no diagnostic group showed a significant preference for these maladaptive patterns.

#### DISCUSSION

Our results indicate a high degree of relationship among several dimensions: acute-chronic, process-reactive, and good-poor premorbid personality. It is, of course, possible that these correlated dimensions have sufficient residual variance to predict differentially various performance or historical criteria. The correlations obtained do not account for more than half the variance in

any instance. A conservative view could justify maintaining these dimensions as distinct until further research with a variety of criteria shows that they can be subsumed under a single heading.

Nonetheless, it seems that at least the good-poor versus process-reactive relationship is functionally meaningful.

The Phillips Scale and the Process-Reactive Scale are highly interrelated, and appear to measure an underlying dimension, which might be labeled "general interpersonal adequacy." Because of its higher interrater reliability, the Phillips Scale is perhaps the preferable research instrument. However, the abbreviated Process-Reactive Scale referred to above may be easier to rate when the patient is guarded or when evaluation is made from incomplete file data. It is interesting that the two raters used much the same cues in making a process-reactive distinction. The failure of some items which have traditionally been associated with one or the other pole of the dimension, e.g., presence of affectivity, is partly a function of the relative homogeneity of the population. Moreover, several items appeared meaningless for this sample. Not one of our subjects showed a "Loss of decency (nudity, public masturbation, etc.)."

The high relationship between the acute-chronic dimension on the one hand, and the process-reactive and good-poor premorbid personality on the other, must be qualified in light of the small number of acute patients involved. Moreover, one must question whether this correlation is meaningful. If process-reactivity is defined in terms of prognosis, then it is obvious that an individual who has been hospitalized for several years is less likely to make a satisfactory recovery than one who has just entered the hospital for the first time.

Of considerable theoretical interest, is the fact that the paranoid-nonparanoid dimension is apparently independent of the "interpersonal adequacy" dimension tapped by the Phillips Scale and the Process-Reactive Scale. Several pieces of evidence attest to this conclusion: the failure of these dimensions to intercorrelate; the significance of differences



between paranoids and nonparanoids on the double alternation task, coupled with the equivalent performance of groups when classified in other ways; and the failure of the "presence of massive paranoia" item to discriminate between process and reactive schizophrenics. The results of the alternation learning confirm the earlier Johannsen (1961) study for paranoids versus nonparanoids. One might speculate that the pertinent variable is one of attention or vigilance. Paranoid individuals are more attentive to the environment than nonparanoids. Whatever differences exist between process and reactives, they apparently are not related to that particular interpersonal function important to success on alternation learning under social feedback conditions.

Whether chronicity is related to alternation learning is not established, since the present group does not show sufficient variability in this respect to settle the question. It is noteworthy, however, that the general level of accuracy (mean number correct = 135.98) is far in excess of that obtained in the Johannsen study (mean number correct = 115.89), which used a group of patients that differed only in the length of total hospitalization.

Our use of interview material rather than case histories (a procedure dictated by the lack of sufficient recorded data) represents a departure in method from the bulk of previous research. The high reliability obtained (at least on the Phillips Scale) is partial justification for this approach. However, before assuming the generality of the correlation between the process-reactive and good-poor personality dimensions, it will be necessary to compare results using both approaches. The authors expect to examine this problem in future research.

#### REFERENCES

- BAXTER, J. C., & BECKER, J. Anxiety and avoidance behavior in schizophrenics in response to parental figures. *J. abnorm. soc. Psychol.*, 1962, **64**, 432-437.
- BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229-236.
- BECKER, W. C. The process-reactive distinction: A key to the problem of schizophrenia? *J. nerv. ment. Dis.*, 1959, **59**, 442-449.
- BLEKE, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1955, **23**, 479-498.
- BLEULER, E. *Dementia praecox*. (Trans. by J. Zinkén.) New York: International Universities Press, 1950.
- BRACEBILL, G. A., & FINE, H. J. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, **52**, 310-313.
- CHAPMAN, L. J., DAY, DOROTHY, & BURSTEIN, A. The process-reactive distinction and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1961, **133**, 383-391.
- COHEN, B. D., SENE, RITA, & HUSTON, P. E. Perceptual accuracy in schizophrenia, depression and neurosis and effects of amytal. *J. abnorm. soc. Psychol.*, 1956, **53**, 363-367.
- DEWOLFE, A. A. The effect of affective tone on the verbal behavior of process and reactive schizophrenics. *J. abnorm. soc. Psychol.*, 1962, **64**, 450-455.
- FARINA, A., & WEBB, W. Premorbid adjustment and subsequent discharge. *J. nerv. ment. Dis.*, 1956, **124**, 612-613.
- FEINBERG, I., & GARMAN, EUGENIA M. Studies of thought disorder in schizophrenia: II. Plausible and implausible errors on a modification of the Progressive Matrices Test. *Arch. gen. Psychiat.*, 1961, **4**, 191-201.
- FINE, H. J., & ZIMET, C. N. Process-reactive schizophrenia and genetic levels of perception. *J. abnorm. soc. Psychol.*, 1959, **59**, 83-86.
- GARMIEZY, N., CLARKE, A. R., & STOCKNER, CAROL. Child rearing attitudes of mothers and fathers as reported by schizophrenic and normal patients. *J. abnorm. soc. Psychol.*, 1961, **63**, 176-182.
- GOLDSTEIN, A. P., & CARR, A. C. The attitudes of the mothers of male catatonic and paranoid schizophrenics toward child behavior. *J. consult. Psychol.*, 1956, **20**, 190.
- GUERTIN, W. H. A factor analytic study of schizophrenic symptoms. *J. consult. Psychol.*, 1952, **16**, 308-312.
- HARRIS, J. G. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, 1957, **25**, 651-671.
- HANEVER, P. H., & IZARD, C. E. Unrealistic self-enhancement in paranoid schizophrenia. *J. consult. Psychol.*, 1962, **26**, 65-68.
- HUSTON, P. E., & SENE, RITA. Psychopathology of schizophrenia and depression. I. Effect of amytal and amphetamine sulfate on level and maintenance of attention. *Amer. J. Psychiat.*, 1952, **109**, 131-136.
- JOHANNSEN, W. J. Responsiveness of schizophrenics and normals to social and nonsocial feedback. Unpublished doctoral dissertation, Univer. of Wisconsin, 1959.
- JOHANNSEN, W. J. Responsiveness of chronic schizophrenics and normals to social and nonsocial feedback. *J. abnorm. soc. Psychol.*, 1961, **62**, 106-113.

- KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, 17, 157-162.
- KING, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenics. *J. abnorm. soc. Psychol.*, 1958, 56, 160-164.
- LONG, R. C. Praise and censure as motivating variables in the motor behavior and learning of schizophrenics. *J. abnorm. soc. Psychol.*, 1961, 63, 283-288.
- LORR, M., MCNAIR, D. M., KLETT, C. J., & LASKY, J. J. Evidence of ten psychotic syndromes. *J. consult. Psychol.*, 1962, 26, 185-189.
- MALLET, J. J. Verbal recall of hostile and neutral thematic contents by schizophrenic and normal subjects. Unpublished doctoral dissertation, Duke Univer., 1956. Cited in E. H. Rodnick & N. Garmezy. An experimental approach to the study of motivation in schizophrenics. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957, 109-184.
- MASON, C. F. Pre-illness intelligence of mental hospital patients. *J. consult. Psychol.*, 1956, 20, 297-300.
- MCDONOUGH, J. M. Critical flicker fusion and the spiral after effect with process and reactive schizophrenics. *J. consult. Psychol.*, 1960, 24, 150-155.
- O'CONNOR, N., & RAWNSLEY, K. Two types of conditioning in psychotics and normals. *J. abnorm. soc. Psychol.*, 1959, 58, 157-161.
- ORGEL, S. A. Differential classification of hebephrenic and paranoid schizophrenics from case material. *J. clin. Psychol.*, 1957, 13, 159-161.
- PHILIP, B. R. Research in the perception of Lissajou figures by psychotics. *Canad. J. Psychol.*, 1953, 7, 115-125.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, 117, 515-525.
- QUERY, JOY M. NEALE. Pre-morbid adjustment and family structure: A comparison of selected rural and urban schizophrenic men. *J. nerv. ment. Dis.*, 1961, 133, 333-338.
- RABIN, A. J., KING, G. F., & EHRLMAN, J. C. Vocabulary performance of short term and long term schizophrenics. *J. abnorm. soc. Psychol.*, 1955, 50, 255-258.
- REISMAN, J. M. Motivational differences between process and reactive schizophrenics. *J. Pers.*, 1960, 28, 12-25.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. of Nebraska Press, 1957, 109-184.
- SAUCER, R. T. A further study of the perception of apparent motion by schizophrenics. *J. consult. Psychol.*, 1958, 22, 256-258.
- SENF, RITA, HUSTON, P. E., & COHEN, B. The use of comic cartoons for the study of social comprehension in schizophrenia. *Amer. J. Psychiat.*, 1956, 113, 45-51.
- SHAKOW, D., & HUSTON, P. E. Studies of motor function in schizophrenia: I. Speed of tapping. *J. gen. Psychol.*, 1936, 15, 63-106.
- SHAKOW, D., & HUSTON, P. E. Studies of motor function in schizophrenia. III. Steadiness. *J. gen. Psychol.*, 1946, 34, 119-126.
- SIEGEL, E. L. Genetic parallels of perceptual structuralization in paranoid schizophrenia: An analysis by means of the Rorschach technique. *J. proj. Tech.*, 1953, 17, 151-161.
- TOPPING, G. G., & O'CONNOR, N. The response of chronic schizophrenics to incentives. *Brit. J. med. Psychol.*, 1960, 33, 211-214.
- WILLIAMS, G. W. A study of the responses of three psychotic groups to a test of suggestibility. *J. gen. Psychol.*, 1932, 7, 302-309.
- WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, 15, 290-302.
- ZAHN, T. P. Acquired and symbolic affective value as determinants of size estimation in schizophrenic and normal subjects. *J. abnorm. soc. Psychol.*, 1959, 58, 39-47.

(Received December 28, 1962)

## THE DIFFERENTIAL PERFORMANCE OF ORALS AND ANALS IN A VERBAL CONDITIONING PARADIGM

EDWIN O. TIMMONS

*Louisiana State University*

CHARLES D. NOBLIN

*University of North Carolina at Greensboro*

The oral character type is said to be dependent and suggestible; anals are said to be obstinate and resistant. Consequently, it was hypothesized that orals would condition better than anals in a verbal conditioning experiment. 24 strong oral or anal undergraduate Ss were selected by the Blacky Test; E did not know the character type of any S. A significant increase in the dependent variable was seen for the orals, while a pronounced drop was found for the anal Ss. The data support the prediction based on Freudian characterological theory.

Through the years the overwhelming majority of concepts used to explain human behavior and to guide the course of psychotherapy have originated from the writings of Freud. Sears (1944), after reviewing the attempts to quantify the tenets and principles of psychoanalytic theory, concluded that experimental psychology had yet to make a major contribution toward stripping away the subjectivism of psychoanalytic concepts. Many consider Sears' judgment too harsh and point to the heuristic value in the experimental analogues of defense mechanisms as evidence (Miller, 1948; Mowrer, 1940; Murray & Berkun, 1955; O'Kelley, 1940; etc.). Further, more or less complete systems of personality and psychotherapy have been offered by Shoben (1949), Dollard and Miller (1950), Mowrer (1950), and Pascal (1959), in which they substituted learning terms for dynamic terms to offer broader understanding and to suggest research approaches to the psychoanalytic system.

The method of verbal operant conditioning, however, has been largely neglected as a tool for directly studying the theoretical concepts which are basic to much of the applied work being done in clinical settings. Obvious advantages of the operant conditioning approach are its empirical, near atheoretical nature and the marked flexibility

it allows the experimenter in choosing both independent and dependent variables.

The present study attempted to utilize operant conditioning techniques to test the predictive efficiency of two of Freud's most venerable personality concepts—the characterological types of *oral* and *anal*. The writings of Freud's followers contain numerous references to the etiology of these character types and to the personality and behaviors expected of them. For instance, Fenichel (1943) and others have suggested that oral character types—being dependent—are highly susceptible to suggestion from authority figures on whose assurance they depend for self-esteem, while anal characters tend to maintain their self-esteem through obstinacy and resistance to authority figures. Consequently, it was hypothesized for this experiment that Oral subjects would condition markedly better than Anal subjects when mild, affirmatory words from an authority figure were used as reinforcing stimuli. A second hypothesis was that more Anals than Orals would be able to verbalize the reinforcement contingency due to the obsessive compulsive or even paranoid method of approaching problems ascribed to anal types by Freudian theory.

### METHOD

The Blacky Test was administered to two undergraduate psychology classes ( $N = 90$ ) from which 15 Oral and 15 Anal subjects were selected for a Taffel-



type verbal conditioning experiment. All the protocols were administered and scored by  $E_2$ ;  $E_1$  conducted all the conditioning portion of the experiment.  $E_1$ , who was the instructor of both classes, did not know which subjects were Orals or Anals until after the conditioning data were completely collected. The scoring was done according to the method developed by Blum (1950). In assigning subjects to oral or anal experimental groups, it was necessary to ignore indicators of other areas of conflict (e.g., masturbation guilt, sexual identification problems, castration anxiety) which are also afforded by the Blacky Test. Blum's method of scoring entails weighing (a) the subject's spontaneous story, (b) answers on a multiple-choice questionnaire, and (c) cartoon preference. Only indicators of orality and anality from these sources were used as criterion measures for this study.

The experimental procedure was standard and simple. The subjects were seen individually, with Orals and Anals randomly intermixed. Each subject was seated in a chair with the experimenter behind and slightly to the left to obviate extraneous reinforcing cues by the experimenter. A series of 120  $5 \times 7$  cards were presented to each subject one at a time. On each at the top were 2 pronouns—one first (I or we), and the other third person (he, she, they). At the bottom of each card was a sentence fragment which, when preceded by either of the pronouns, made a complete sentence. This deck of cards has been refined, and each type of pronoun is equally represented and randomized as to position within each of the 12 blocks of 10 cards.

The first 30 cards were used to establish an operant level for each subject's choosing the selected class of pronoun. (For this experiment, first person pronouns were designated as the "correct" response.) During the next 60 trials mild, affirmatory utterances such as "Um-hmm," "That's fine," "O.K.," "Good,"

etc., were administered following correct responses on a 75% variable ratio schedule. The schedule was employed to help minimize the number of subjects who might verbalize the reinforcement contingency. The last 30 cards constituted the extinction phase in which no reinforcements were given for either choice.

After each subject's responses had been collected and recorded, he was asked the following three questions in an attempt to assess awareness: (a) Now, what do you think this experiment was all about; what was the general idea? (b) Did you notice MY saying or doing anything? (answer) Why was I doing that? (c) Did you notice any change in YOUR responses from the first of the cards to the last? (answer) How did you change? (answer) Why did you change?

Any subject who verbalized the experimental principle was omitted from the analysis of the conditioning data.

## RESULTS

Figure 1 shows the conditioning data for each experimental group plotted in blocks of 30 trials. The curve generated by the Oral subjects shows the typical verbal conditioning curve: a level of response near chance during the operant phase, a continuing increase in the dependent variable throughout the treatment phase, and a drop back to near-operant level during extinction. Friedman's nonparametric analysis of variance for correlated data showed the overall effect of treatment to be significant beyond the .01 level. Wilcoxon's matched-pair  $T$  tests applied to the points along the curve reveal that the source of the significance is due to the difference between the operant level and the second half of treatment level ( $p < .01$ ), and to the drop from the second half of treatment to the extinction level ( $p < .05$ ). The other three comparisons failed to reveal significant differences.

A marked contrast is noted when the performance of the Anal subjects is plotted. The Anal subjects started at a high level (56%) in choosing the personal pronouns during operant, then dropped nearer chance (53%) during the first 30 treatment trials, and then dropped to a below chance frequency (48.6%) during the second 30 treatment trials. When the experimenter discontinued reinforcement during extinction, the fre-

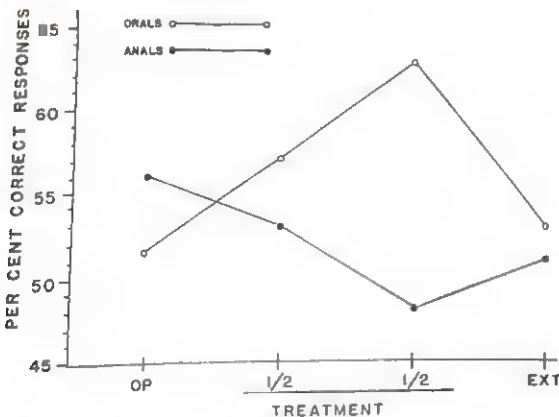


FIG. 1. Performance curves in Taffel-type conditioning situation for 15 subjects rated as "Anals" and 15 subjects rated as "Orals" on Blacky Test. (Data plotted in four blocks of 30 trials each.)

quency of the dependent variable rose to near chance again (51.5%). The curve approached, but did not reach, the accepted significance level.

Direct comparison *between* the two groups at the four points along the curves was afforded by the Mann-Whitney nonparametric *T* test for independent data. The Oral and Anal groups' operant levels, the extinction levels, and the first half of treatment for each all failed to yield significance; but the difference between the Orals' 63% and the Anals' 48% during the second half of the treatment phase was significant beyond the .05 level. All *p* levels reported are two-tailed.

### DISCUSSION

Hypothesis 1 was supported by the data, for two groups of students treated exactly alike and differentiated only by Oral and Anal scores on the Blacky Test showed strikingly different performance on a standard verbal conditioning problem.

Hypothesis 2 was not supported. It had been predicted that the suspicious, precise approach to problems said to be associated with the Anal character would lead to more Anals than Orals verbalizing the idea of the experiment. Actually, six subjects verbalized the experimental concept—three Orals, three Anals.

A post hoc re-analysis of the data was made to see how consistently one could have predicted subjects' Oral or Anal rating on the Blacky Test from the conditioning data alone. Following the method used by Matarazzo et al. (1960), each subject was tallied in a  $2 \times 2$  table as having "gone up" from his mean number of personal pronouns given during operant to his mean number given during the treatment phase, or having "gone down" from operant to treatment phase. Oral and Anal subjects were tallied separately. The resulting table showed that of the 15 Oral subjects, 12 showed an increase and 3 stayed the same; none went down. Of the 15 Anal subjects, 4 increased from operant to treatment phases and the remaining 11

showed a decrease over the same comparison. A chi square comparison of these data yielded a *p* value considerably less than .01, reflecting the consistency of the data.

Since the execution of this experiment, closely related findings have been reported by Cairns and Lewis (1962). These investigators used the Edwards Personal Preference Schedule to divide subjects into High and Low Dependency groups, then placed the entire group in a verbal conditioning situation very similar to ours. The two sets of findings were almost identical, for their Low Dependency subjects behaved very like our Anal subjects in showing a sharp decrement in emission of the reinforced verbal class; their High Dependency subjects showed a conditioning effect, though not as pronounced as our Orals. Further, an extension of the Oral-Anal predictor type experiment has just been completed by one of our students which replicated the findings of both this and the Cairns and Lewis study. It would appear that verbal operant conditioning can be a useful tool in investigating clinical concepts.

### REFERENCES

- BLUM, G. S. *The Blacky pictures: A technique for the exploration of personality dynamics* (manual). New York: The Psychological Corporation, 1950.
- CAIRNS, R. B., & LEWIS, M. Dependency and the reinforcement value of a verbal stimulus. *J. consult. Psychol.*, 1962, 26, 1-7.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1943.
- MATARAZZO, J. D., SASLOW, G., & PAREIS, E. N. Verbal conditioning of two response classes: Some methodological consideration. *J. abnorm. soc. Psychol.*, 1960, 61, 190-206.
- MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 155-178.
- MOWRER, O. H. An experimental analogue of regression with incidental observations on reaction formation. *J. abnorm. soc. Psychol.*, 1940, 35, 56-87.
- MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.

- MURRAY, E. J., & BERKUN, W. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, 51, 47-56.
- O'KELLEY, L. I. An experimental study of regression, I. and II. *J. comp. Psychol.*, 1940, 30, 41-53 and 55-95.
- PASCAL, G. R. *Behavioral change in the clinic—A systematic approach*. New York: Grune & Stratton, 1959.
- SEARS, R. R. Experimental analysis of psychoanalytic phenomena. In J. McV. Hunt (Ed.). *Personality and the behavior disorders*. New York: Ronald Press, 1944.
- SHOBEN, E. J. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, 46, 366-393.
- ZELLER, A. F. An experimental analogue of regression. I. Historical summary. *Psychol. Bull.*, 1950, 46, 39-51.

(Received October 25, 1962)



## FREQUENCY HYPOTHESIS AND CONTENT ANALYSIS OF PROJECTIVE TECHNIQUES

JOHN WALLACE<sup>1</sup> AND LEE SECHREST<sup>2</sup>

*Northwestern University*

An attempt was made to evaluate the frequency hypothesis of projective theory by constructing a multitrait-multimethod matrix for 4 traits and 5 methods. Traits measured were somatic concern, hostility, achievement concern, and religiosity. Methods were self-description, peer reputation, content Rorschach, Incomplete Sentences Blank, and Thematic Apperception Test. Additional measures were number of visits to health services, medical symptom check list, and scholastic average. Ss were student nurses. The frequency hypothesis received support only for specific traits as measured by specific methods. Self-description, peer reputation, and ISB yielded the most promising convergent and discriminant validities. Some evidence supported the validities of ISB and TAT for measurement of religiosity while the Rorschach had some convergent validity for measurement of hostility. Self-description and peer reputation measures yielded consistently higher correlation than any other combination of methods.

Lindzey (1961) has made explicit a number of hypotheses which underlie the use of projective techniques. The research reported here concerns one of these hypotheses which we have termed the "frequency hypothesis." The frequency hypothesis states that "Those responses elicited or produced under a variety of different stimulus conditions are particularly likely to mirror important aspects of the subject [p. 149]." Levy (1963) in a discussion of the logic and assumptions of psychological interpretation has elaborated the frequency hypothesis in learning theory terms:

There would seem to be some support for this assumption from the vast amount of research in learning theory, whereby it is generally found that the frequency of occurrence of a response increases (or alternatively, its probability of occurrence increases) with increased training and reinforcement [p. 51].

Lacking knowledge of a person's past history of reinforcement, the clinician may attempt to predict future test or extra-test behavior by focusing upon frequency of occurrence of certain responses during testing. Frequency

of response may be construed, then, as one potentially useful measure of response strength (Auld, 1954).

In the present research, an attempt was made to investigate the usefulness of simple frequency counts in conjunction with content analysis of projective techniques. Frequency counts of specific content have yielded promising results in several studies (e.g., Fisher & Cleveland, 1958; Harris, 1960). Moreover, concentration on content rather than determinants and the use of simple linear rather than configural methods of combining data hold considerable promise for high scorer agreement. Obviously, appreciable agreement among scorers is a necessary minimal requirement for adequate measurement.

The traits measured were hostility, somatic concern, achievement concern, and religious concern. The methods of measurement employed were the Rorschach, selected Thematic Apperception Test cards (TAT), Rotter's Incomplete Sentences Blank (ISB), self-description, peer reputations, visits to the student health service (SHV), first year grade point average (GPA), and a typical medical symptom check list (MSC).

In selecting traits to be measured, a deliberate attempt was made to select at least one clearly socially acceptable trait as well as one socially disapproved trait. It appeared reasonable, for this sample of subjects, to

<sup>1</sup> This study was completed while the first author was a USPHS predoctoral research fellow at Northwestern University.

<sup>2</sup> The writers wish to thank Bernard Schiff, Judy Hunt, Fred Levine, and Virginia Smith who assisted in the judgment of responses and analysis of data.

consider religious concern as high in social acceptability and hostility as low in social acceptability. An attempt was made to select methods which vary in degree of structure. For example, the Rorschach represents the most unstructured and indirect of the methods of measurement while self-description represents the most structured and direct. In order to assess the relative validity of the projective techniques, self-description was included to provide a base line against which to evaluate the more unstructured techniques. Obviously, if projective techniques are ultimately to prove valuable, they should provide more accurate information than can be obtained by the simple and far less costly technique of asking the subject to describe himself. In the present investigation, peer reputations were considered a principal, although admittedly imperfect, criterion.

In order to assess the discriminant validity of the methods of measurement, a multitrait-multimethod matrix was constructed (Campbell & Fiske, 1959). In brief, the multitrait-multimethod matrix provides the means through which the extent of true (trait) variance as well as unwanted methods of measurement can be assessed.

## METHOD

### Procedure

The subjects consisted of 72 freshmen nursing students enrolled in an introductory psychology course at a large midwestern university. All of the testing was conducted in groups in two separate testing sessions separated by a time interval of 3 days. The most direct measures, self-description, peer reputation, and check list, were the last measures gathered. Standard instructions for group administration of the Rorschach, TAT, and ISB were read. In the case of the TAT and Rorschach, the stimuli were projected on a large white screen and the subjects were instructed to write their responses in previously prepared booklets. An exposure time of 5 minutes was allowed for each TAT card and each Rorschach card. All of the Rorschach cards were presented. In the case of the TAT, the following cards were presented: 1, 3GF, 12, 7GF, 3BM, 8GF.

The self-description and peer reputation ratings consisted of simple rating scales of 28 items, 7 items for each trait. Subjects were asked to distribute their ratings of themselves and their roommates over a 9-point scale. Sample items from the rating scales were as follows: *Somatic concern*: "has vague aches and pains," "concerned about personal health." *Hostility*: "gets mad at others," "criticizes others."

*Achievement concern*: "concerned about grades," "is dissatisfied with a C on exams." *Religious concern*: "concerned with Christian values," "interested in church matters." At the time the ratings were gathered, the subjects had been living together for at least 6 months in small apartments which housed a maximum of four persons. In nearly all cases, each girl rated her three roommates and was rated by her three roommates. Considering the nature of the living arrangements, it seems reasonable to conclude that the subjects knew one another quite well.

### Scoring Systems

*Rorschach*. As mentioned above, no attempt was made to elicit the determinants of the responses. All Rorschach scoring was accomplished on the content of the responses. Only the most obvious percepts were scored as positive for the trait under investigation. Thus, for Somatic concern all anatomical responses were considered relevant (Klopfer, Ainsworth, Klopfer, & Holt, 1954, p. 383). Percepts such as bones, X rays, viscera, and internal organs, as well as diseases such as cancer and tubercular lung, were all scored as positive for somatic concern. Blood with no mention of violence was also scored for somatic concern. In the case of Hostility, all expressions of violence, human or animal, were scored as relevant (Klopfer, 1954, pp. 381, 383). In addition, all implements of violence were also scored (Holtzman, Thorp, Swartz, & Herron, 1961). Thus, percepts such as gun, sword, people fighting, animals devouring one another, warriors, battle, etc., were scored as relevant for hostility. Rorschach Religious concern was scored for percepts such as cross, Jesus, monks, church, temple, hands in prayer, etc. It was not possible to score Rorschach content for Achievement concern.

*Incomplete Sentence Blank*. As with Rorschach, attention was again focused upon the content of the responses in the scoring of the ISB. For Somatic concern, all responses suggesting concern with one's physical well being were scored as positive. Thus, responses which mentioned actual physical illness of the subject or a person known to the subject were scored. Expressions of fatigue were also considered indicative of somatic concern. A similar scoring method for ISB somatic concern was used by Sechrest and Hemphill (1954). For Hostility, all responses which indicated interpersonal hostile-aggressiveness were scored. Thus, "I hate my biology instructor" was scored as positive while, "I hate to study biology" was not. Responses which suggested Achievement concern were as follows: "My greatest fear is flunking out," "Exams really upset me," "I really regret that I didn't study enough." In the case of Religious concern, responses which obviously dealt with religion were scored. For example, "at bedtime, I read my Bible and pray," "I want to serve God with all my heart and soul" were scored as positive for religious concern.

*Thematic Apperception Test*. The scoring of the TAT differed somewhat from that of the Rorschach and the ISB. No attempt was made to judge the



overall content, thema, or story ending since TAT stories are frequently so long and complex that the assignment of a simple score can prove difficult. Rather, simple word counts were conducted. The number of words thought indicative of the four traits were recorded. Words which were scored as relevant for Somatic concern were as follows: "sick," "cancer," "fever," "illness," "heart attack," and "disease." For Hostility, words such as the following were scored: "murder," "destroy," "shoot," "fight," "argue," and "strike." For Achievement concern, the following serve as examples: "study," "work," "ambition," "learn," "school," and "master." In the case of Religious concern, words such as the following were scored: "Jesus," "praying," "sin," "God," "blessing," "last rites," "priest," and "minister."

**Scorer Agreement.** Since the resulting responses amounted to an almost overwhelming mass of data, the initial judgments of the responses were conducted by the first author, who sorted all stimuli into relevant and irrelevant categories for the traits under investigation. Prior to the experimenter's judgments, all of the Rorschach and ISB responses were typed by another person on individual  $3 \times 5$  white index cards, which were then shuffled. The judgment procedure was truly "blind" in that the experimenter had no information other than the isolated single responses which were to be judged. After the experimenter had made his initial judgments, 100 randomly selected irrelevant responses were added to these Rorschach and ISB responses judged relevant. This new set of responses, relevant and irrelevant, were presented to a second judge together with instruction for making the judgments. Three different judges were employed, one for each of the three projective measures. In the case of the ISB, the second judge categorized each response as relevant to one of the four traits or as irrelevant. Thus, chance agreement between the experimenter and the second judge was  $\frac{1}{4}$  or 20%. For the ISB, the actual agreement between the experimenter and the second judge was 90%. In the case of the Rorschach, judgments were made of three traits and one irrelevant category with chance agreement of  $\frac{1}{4}$  or 25%. An actual agreement value of 96% was obtained between the experimenter and the second judge for the Rorschach. It should be noted that a spuriously high agreement value arising out of an extremely disproportionate number of irrelevant responses was not possible by this method.

For the TAT, scorer agreement was demonstrated as follows. All of the TAT stories were typed on individual sheets of paper and shuffled. Again, the experimenter made his judgments with no information other than individual TAT stories. After the experimenter made his judgments, 85 randomly selected stories were obtained from the total pool of TAT stories. These 85 protocols were presented to a second judge with appropriate instructions. The correlations between the experimenter and the second judge for the TAT were as follows: somatic concern, .93, hostility, .92, achievement concern, .87, and religious concern, .94.

In general, it seems justified to conclude that the scoring systems produced considerable agreement. Considering the fact that attention was focused upon the most obvious attribute of the response, the content, the high agreement values which were obtained are not surprising.

### *Analysis of the Data*

Prior to correlational analysis, the raw scores from several of the methods were converted to indices. Since neither the TAT nor the Rorschach controls the number of responses a subject may make to a single stimulus, it was necessary to convert each trait score into a ratio score. This was accomplished by dividing each trait score by the total number of responses. In this manner, a possible spurious correlation between trait scores and total responses was made considerably less likely. For the peer reputational scores, mean peer reputational scores were computed and employed in the correlational analysis. All of the possible inter-correlations among the 22 variables were obtained, resulting in a matrix that consisted of 231 Pearson product-moment correlation coefficients.

### RESULTS

Table 1 presents the multitrait-multimethod matrix obtained by intercorrelating the data obtained for the four traits by the seven methods of measurement. Utilizing peer reputational ratings as a principal criterion, it is obvious that the highest validity coefficients in the entire matrix involve self-description and peer reputation. All of the peer reputational self-description diagonal values are significant beyond the .01 level of significance. Moreover, the validity diagonals are clearly higher than the off diagonal heterotrait-heteromethod values as well as the values contained in each of the respective monomethod-heterotrait groups of correlations. Thus, both methods, peer reputation and self-description, meet the criteria for reasonably good convergent and discriminant validity.

It is encouraging to note that both self-description and peer reputation yield significant relationships with behavioral measures. For achievement concern, both of the measures correlate significantly with scholastic average at the close of the academic year. For somatic concern, peer reputation correlates significantly with number of visits to the infirmary while self-description does not. However, on a second self-description



TABLE 1  
MULTITRAIT-MULTIMETHOD MATRIX

Self-description	Self-description				Rorschach			TAT			ISB			Peer reputation				SHV MSC GPA					
	S <sub>1</sub>	R <sub>1</sub>	H <sub>1</sub>	A <sub>1</sub>	S <sub>2</sub>	R <sub>2</sub>	H <sub>2</sub>	S <sub>3</sub>	R <sub>3</sub>	H <sub>3</sub>	A <sub>3</sub>	S <sub>4</sub>	R <sub>4</sub>	H <sub>4</sub>	A <sub>4</sub>	S <sub>5</sub>	R <sub>5</sub>	H <sub>5</sub>	A <sub>5</sub>	S <sub>6</sub>	S <sub>7</sub>	A <sub>6</sub>	
Self-description																							
Somatic S <sub>1</sub>																							
Religion R <sub>1</sub>																							
Hostility H <sub>1</sub>																							
Achieve A <sub>1</sub>																							
Rorschach																							
Somatic S <sub>2</sub>																							
Religion R <sub>2</sub>																							
Hostility H <sub>2</sub>																							
TAT																							
Somatic S <sub>3</sub>																							
Religion R <sub>3</sub>																							
Hostility H <sub>3</sub>																							
Achieve A <sub>3</sub>																							
ISB																							
Somatic S <sub>4</sub>																							
Religion R <sub>4</sub>																							
Hostility H <sub>4</sub>																							
Achieve A <sub>4</sub>																							
Peer reputation																							
Somatic S <sub>5</sub>																							
Religion R <sub>5</sub>																							
Hostility H <sub>5</sub>																							
Achieve A <sub>5</sub>																							
SHV																							
Somatic S <sub>6</sub>																							
MSC																							
GPA																							

Note.—N = 72;  $r > .23$ ,  $p < .05$ ;  $r > .30$ ,  $p < .01$ .

measure, number of medical symptoms checked, a significant relationship with visits to the infirmary was found.

Among the projective techniques, it is apparent that the ISB yielded the most encouraging validity values. Inspection of the ISB validity diagonals reveals five significant validity coefficients. With the exception of an obtained significant negative correlation between ISB somatic concern and peer reputation for achievement, the discriminant validity of the ISB appears acceptable. The results for the Rorschach and the TAT are somewhat more discouraging. The Rorschach yielded but one significant validity value and the TAT only two. Moreover, the discriminant validity of both methods of measurement was found wanting in that both yielded a number of heterotrait-heteromethod values as high or higher than the obtained validity values. In addition, the obtained monomethod-heterotrait correlations for the TAT were in two instances approximately equivalent to the obtained two significant validity diagonals. In summary, self-description and peer reputation appeared to be acceptable methods of measurement for all of the traits measured. The ISB yielded significant relationships for hostility, religious concern, and somatic concern. Moreover, the discriminant validity of the ISB appeared acceptable. The TAT, scored by the present method, may prove of some use as a measure of a socially acceptable trait, religious concern. The Rorschach, on the other hand, yielded one significant validity value for a socially unacceptable trait, hostility. However, both methods of scoring the Rorschach and the TAT leave much to be desired in terms of discriminant validity.

#### DISCUSSION

Obviously, the present research does not answer the question of the validity of the Rorschach, ISB, and TAT per se. Such a general question is meaningless. An attempt was made to evaluate a specific hypothesis, the frequency hypothesis, in the context of specific scoring systems based upon content analysis. While it is apparent that the present research does not address itself to the general

question of the validity of projective techniques, it should be equally clear that the investigation does not provide a final answer as to the usefulness of the frequency hypothesis. However, the frequency hypothesis does appear tenable when used in conjunction with content analysis of projective techniques for some traits and some methods.

Of considerable interest in the present investigation is both the convergence as well as the lack of convergence between projective measures and more public measures of personality. With the exception of the Rorschach, it is clear that all of the measures converged for the socially acceptable trait, religious concern. Thus, for a socially acceptable trait, frequency appears to be a very useful measure of response strength. However, with the exception of Rorschach content, measurement of a socially unacceptable trait, hostility, resulted in an appreciable decrease in validity. The fact that Rorschach content for hostility correlated significantly only with self-description is indeed surprising. However, it may well be the case that the content of the Rorschach response is closely related to the other rather accessible verbal responses of the subject while the determinant of the response is not. Clearly, the present research does not provide an answer to this question.

Since the present research is concerned primarily with the degree of convergence among methods varying in degree of structure, it is important to consider possible sources of method variance. Of rather critical concern is the amount of "hidden" method variance which might exist between methods which, at an obvious level, appear to be independent. While it might be thought that self-description and peer reputation are independent, it seems likely that they share two important methods factors. First, self-description and peer reputation measures usually involve a common descriptive vocabulary, e.g., when the subjects and their peers are asked to make ratings on "hostility." Such commonality of descriptive terminology does not exist between such methods as the Rorschach and TAT. Second, self-description and peer reputation measures are not independent from the standpoint of method by

reason of the fact that each measure very likely reflects something of the other. They very probably develop in a reciprocal way. Attitudes toward the self develop, in part, out of the reactions of others to oneself. And our reputations are probably based in part on our descriptions of ourselves to others.

However, it would very probably be erroneous to conclude that peer reputations are to be dismissed as "nothing but" one or the other of the above notions. Certainly, among reasonably sophisticated and intelligent populations, impressions about others are not formed solely on the basis of what others say or show directly about themselves. Marlowe (1962) has shown that fraternity men who paint an exaggeratedly favorable picture of themselves are less well liked than those who admit to what would seem to be flaws of personality and character. In such a case we might conclude that the peer reputations are indeed insightful and reflect something other than persons' verbalizations about themselves.

In some respects it may be helpful to regard a given behavior from several vantage points. First, one may think in a rather abstract or theoretical sense of the underlying "behavioral disposition" (Campbell, 1963) which will at some point become overt and thus observable by others. However, in the process of being translated into an overt response, a behavioral disposition will be filtered through the cognitive and other systems of the behavior, perhaps being completely altered in the process. Thus, for example, an underlying disposition of hostility toward other persons may, because of a past history of negative reinforcement for such behavior, be inhibited. When the individual's past history of negative reinforcement for hostile behavior is known (Lesser, 1957) and when expectancies of punishment for aggressive behavior are present (Mussen & Naylor, 1954), the relationship between projective responses and overt behavior is much clearer.

In addition to modification or inhibition of underlying behavioral dispositions, one must consider the *intent* of the behavior, i.e.,

what effect the behavior means to produce. However, it would be clearly incorrect to assume a one to one relationship between the intent of the behavior and the *impact* of his behavior upon others. In other words, the intent of the behavior may not be reflected accurately in the response which he arouses in others. Thus, a response which the behavior intends as hostile may in actuality be so inept or may strike such an insensitive environment that its intended hostility is missed.

Obviously, then, a given measure may tap into the behavior eventuation process at one of three points: the underlying behavioral disposition, the intent of the behavior, and the impact of behavior upon the environment. It appears reasonable to assume that self-description has a strong component of intent, whereas peer reputations should be more nearly a measure of impact. Moreover, most theorists and clinicians would be inclined to agree that projective measures *should* provide some measurement of the "pure" or essential behavioral disposition underlying the whole process or some measurement of the transforming operations that may intervene between disposition and intent.

In the measurement of a socially acceptable trait, one would expect to find considerable congruence among disposition, intent, and impact. Indeed, in the present investigation, measurement of religious concern yielded the highest validity coefficients for all measures with the exception of the Rorschach. The failure of the Rorschach to yield significant correlations for religious concern may well be a function of the stimuli. Across all Rorschach cards, the "pull" for religious percepts may be, in actuality, so minimal as to preclude frequent occurrence of such percepts, even by the religious zealot.

In the measurement of socially disapproved traits, one would hardly expect to find appreciable congruence among disposition, intent, and impact. It is interesting to note that the projective measures, with the exception of the trait religious concern, yielded significant correlations only with self description. For example, the ISB which correlated significantly with self-description for somatic



concern, religious concern, and hostility yielded nonsignificant relationships with peer reputational measures of somatic concern and hostility. On the other hand, the correlations representing the relationships between ISB self-description and ISB peer reputation are of approximately equal magnitude and clearly significant.

Of course, it cannot be maintained that the three response components we have described are in practice ever separable, for it is very clear that behavioral disposition and intent are often overlapping, if not identical, even for undesirable traits. Similarly, response impact may reflect not only the underlying disposition and the intent of the actor, but various transforming and defensive operations as well. It would appear unwarranted to assume that the intent of one's behavior is never correctly perceived and appropriately responded to by one's peers, but it would appear equally unwarranted to assume that intent is always clearly reflected in peer reputations. It is of definite interest that the relationship between self-description and peer reputation holds for both the seemingly desirable and ordinarily transparent trait of religiosity and for the less acceptable trait of hostility.

The fact that projective measures of hostility and somatic concern correlate significantly only with self-description suggests that for traits other than socially acceptable ones, something is lost between the response and its impact. One hypothesis that might merit further investigation is that projective measures tap an aspect of a behavioral dis-

position of which the actor may be at least partially aware but which is either so transformed in overt response or is so subtle in its manifestation that its impact is nil.

#### REFERENCES

- AULD, F. Contributions of behavior theory to projective techniques. *J. proj. Tech.*, 1954, 18, 421-426.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, 56, 81-105.
- FISHER, S., & CLEVELAND, S. E. *Body image and personality*. Princeton: D. Van Nostrand, 1958.
- HARRIS, J. G. Validity: The search for a constant in a universe of variables. In M. A. Rickers-Ovsiankina (Ed.), *Rorschach psychology*. New York: Wiley, 1960. Pp. 380-436.
- HOLTZMAN, W. H., THORP, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality*. Austin: Univ. Texas Press, 1961.
- KLOPPER, B., AINSWORTH, M. D., KLOPPER, W. G., & HOLT, R. R. *Developments in the Rorschach technique*. New York: World Book, 1954.
- LESSER, G. S. The relationship between overt and fantasy aggression as a function of maternal response to aggression. *J. abnorm. soc. Psychol.*, 1957, 55, 218-221.
- LEVY, L. H. *Psychological interpretation*. New York: Holt, Rinehart & Winston, 1963.
- LINDZEY, G. *Projective techniques and cross-cultural research*. New York: Appleton-Century-Crofts, 1961.
- MARLOWE, D., & BANK, S. Need approval and need affiliation as determinants of "likeability." Unpublished research, Dartmouth College, 1962.
- MUSSEN, P. H., & NAYLOR, H. K. The relationships between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, 49, 235-240.
- SECHREST, L., & HEMPHILL, J. K. Motivational variables in the assuming of combat obligation. *J. consult. Psychol.*, 1954, 18, 113-118.

(Received December 11, 1962)

# THE RELATIONSHIP OF EXPECTANCY OF REWARD TO ACHIEVEMENT PERFORMANCE ON AN ARITHMETIC AND THEMATIC TEST<sup>1</sup>

BERNARD I. MURSTEIN<sup>2</sup>

*Connecticut College*

Atkinson's theory of motivation was tested with regard to a thematic test scaled for achievement value and a simple arithmetic task. Three randomly selected groups, of approximately 20 college men each, vied for a \$1.25 cash prize, their objective probability of success being .1, .5, and .9. After the achievement arousing instructions they received a thematic test followed by the arithmetic one. Hypothesis 1 predicted the greater manifestation of  $\pi$ -Ach for the .5 group over the others. Hypothesis 2 predicted a similar superiority for the arithmetic task. The results largely supported Hypothesis 1, but not Hypothesis 2. The stimulus properties were found to greatly influence the response, the medium-achievement structured cards being the most differentiating.

In recent years Atkinson (1958) has proposed a theory to account for the expression of motivation to achieve under a variety of circumstances. The theory states that the response is a function of the product of the motives operating on a subject, his expectancy of goal attainment through making this response (subjective probability of success,  $P_s$ ), and incentive ( $1 - P_s$ ). The equation representing incentive follows the logic that the goal which is the least likely to be attained possesses the greatest attraction. The approach gradient of motivation to achieve is represented by

$$\text{Motivation to Achieve} = M_a \times P_s \times I_s$$

where

$M_a$  = strength of motive to achieve

$P_s$  = subjective probability of success

$I_s$  = the incentive value for achievement  
=  $1 - P_s$ .

An avoidance gradient such as "fear of failure," where measured, would be represented by the following equation:

$$\text{Motivation to Avoid Failure} = M_f \times P_f \times I_f$$

<sup>1</sup> The study was carried out through the support of a Public Health Service Grant (M-4698) from the National Institute of Mental Health.

<sup>2</sup> The author wishes to express his thanks to Martyn Spitzer for his aid in carrying out the study.

where

$M_f$  = motive to avoid failure

$P_f$  = probability of failure

$I_f$  = the incentive for avoiding failure =  
-  $P_s$ .

The resultant motivation equals the net sum of adding the two gradients.

$P_s$  may be differentiated from the objective probability of success  $P_o$  since a person's estimate of his chance of succeeding at a task may depart from the objective probability of success. If, for example, the top 10 out of 20 students are to receive a prize for their performance, some students will perceive their chances as better than one in two, while others will not. Earlier work (Easter & Murstein, in press) has indicated that for a group of randomly selected college men, the median value of estimated subjective probability of success and the objective probability of success are approximately equal.

If through the process of randomly selected subjects, motive strength which is considered to be a stable personality characteristic by Atkinson, is held constant as well as motivation to avoid failure, then Motivation to Achieve becomes a function of  $P_o \times 1 - P_o$ . This product is at a maximum when the probability of success is .5, since  $.5 \times (1 - .5) = .25$ , and is lower for any other value of  $P_o$ , the distribution taking the shape of a parabolic curve,

Previous studies by the author and colleagues have supported the effectiveness of the theory regarding the prediction of performance of children on simple arithmetic tasks, but not on the projection of achievement or thematic cards (Murstein & Collier, 1962). With college men instructed to make up "creative stories," which were then scored for *n*-Ach, only partial support of the theory resulted (Easter & Murstein, in press).

These studies, however, did not fully control key factors which might be expected to influence the character of the response. These factors included (a) the use of combined sexes which masked possible sex differences, and (b) the failure to adequately determine the impact of the thematic stimulus upon the response. The earlier experiments were essentially repeated in the present study, therefore, controlling for these two variables in that only college men were employed and the stimulus pull of the thematic cards was measured with regard to judged achievement value prior to the experiment.

The hypotheses were:

1. A group of randomly selected college men having a .5 expectancy of receiving a cash prize of \$1.25 contingent upon the excellence of their performance will project more achievement imagery than (a) a group having a .1 expectancy of success, and (b) will also project more achievement imagery than a group having a .9 success expectation.
2. The .5 group will likewise under similar conditions attempt more arithmetic examples than either the .1 or .9 groups.

### PROCEDURE

Fifty-six college men from two introductory psychology classes were randomly divided into three sections. The first section received the following instructions from the examiner:

We are conducting an experiment in which we want to see what you can do on arithmetic problems. When I say go, you start and keep working until I tell you to stop. There are quite a few examples so it is unlikely that you can finish all of them. Those persons finishing in the top 10% in terms of number correct will receive a cash prize of \$1.25. There are approximately 20 persons here, so that means the two highest scores will receive the prize. (For the other two groups,

the winning percentage was set at 50% and 90%, respectively).

After all questions had been answered, *E*<sup>3</sup> continued, saying:

Due to the shortage of time remaining and the length of the arithmetic task, we will not give out the arithmetic sheets today, but will do so at your next meeting in two days. Today, we would like to have you participate in a normative study of college responses to the TAT. This project has been going on since last term and is designed to obtain an idea of how college students respond to the various TAT cards.

At this point the usual McClelland instructions for writing TAT stories were given, including the projection of the cards for 20 seconds and the allotment of 4 minutes for the writing of the story. Nine cards were used, seven TAT cards, and two which were selected from the list provided by Atkinson in his book (Atkinson, 1958a, pp. 832-834). These cards had been selected from 38 cards scaled by the Likert technique for amount of achievement depicted, on the basis of differentiating between high and low judged achievement scores for all of the cards, and represented the entire range of the continuum. The cards had been divided into three subsets of three cards each: High (pioneer chopping down tree—Atkinson no. 98—2, 17BM), Medium (8GF, 14, boy at desk with open book in front of him—Atkinson no 8), and Low (20, 3BM, 13B). They were presented in a single order randomly arrived at, except that every three cards had to contain one card from each of the subsets. The order of administration was: card showing pioneer chopping down tree, 8GF, 20, 3BM, 14, 2, 17BM, boy in checked shirt at a desk with an open book in front of him, 13B.

On the following meeting of the class, the experimenter repeated the instructions, the arithmetic examples were given, and a time period of 20 minutes allotted for the 600 arithmetic problems. These consisted of simple arithmetic or subtraction examples of the type  $33 + 71 = ?$  and  $42 - 29 = ?$  Only one of the students finished all of the examples, this student doing so just as the time period elapsed.

The accuracy of the scorer in scoring the stories was verified prior to the experiment by comparing his scoring of a series of stories listed in Atkinson (1958a) with the scores given the stories in the

<sup>3</sup> Because, as Birney (1958) has shown, the person giving the test whether student or professor may profoundly affect the amount of achievement earned, certain precautions were taken. Both the author and his assistant, a senior college student, essayed a relaxed but not informal role, being neatly attired in white shirts, but wearing sweaters. The role of the administrators was explained as "participating in the research project." As far as was known, neither of the administrators was known to the students.



book by experts. The interscorer reliability of .89 was deemed sufficiently high to assure adequate reliability of scoring for the present study.

### RESULTS

To test the first hypothesis, the means and standard deviations of *n*-Ach were obtained for each of the groups and are shown in Table 1 along with the means and standard deviations of the low, medium and high subsets. The *t* values are shown in Table 2 for the group comparisons. This table shows that the .5 group as predicted manifested significantly more *n*-Ach than the .1 group. A nonsignificant trend is evident when the .5 group is compared to the .9 group, with the .5 group manifesting more *n*-Ach. The importance of the stimulus structure is further indicated in Table 2 by the fact that the medium structured cards are shown to be responsible for the superiority of the .5 group.

The second hypothesis was tested by computing the mean number of arithmetic examples attempted and running *t* tests between the groups as shown in Table 3. No differences were found for any of the comparisons, and the hypothesis is thus not supported.

Further analysis was undertaken to see whether the number of arithmetic examples, correctly solved, differentiated the groups, but *t* tests showed no significant differences.

An unanticipated event which occurred during the course of the experiment provided an indirect test of the assumption that achievement imagery and number of arithmetic examples attempted were valid indices of achievement motivation. Nine students of the .1 group did not show up for the second

TABLE 1

MEANS AND STANDARD DEVIATIONS OF *n*-ACH SCORES FOR THE TOTAL, LOW, MEDIUM, AND HIGH SET OF THEMATIC CARDS

Group	Total		Low		Medium		High	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
.1	-.28	1.42	-.81	.76	-.66	.82	.61	1.89
.5	.18	1.77	-.61	1.04	.27	1.90	.88	1.90
.9	-.14	1.60	-.77	.74	-.67	.53	1.02	2.21

TABLE 2

*t* VALUES FOR THE DIFFERENCES IN *n*-ACH BETWEEN THE GROUPS

Groups	Total	Low	Medium	High
.9 vs. .5	1.67	.90	3.42**	.35
.5 vs. .1	2.60**	1.15	3.26**	.76
.9 vs. .1	.98	.28	.07	1.07

\*  $p < .05$ .

\*\*  $p < .01$ .

session in which the arithmetic task was administered. These persons were subsequently given the task at the next session, however, and are included in the .1 group scores. Since not a single one of the subjects of the other groups missed the second session, these subjects' absence is hardly attributable to chance. The only apparent reason for their absence seems to be the low probability of their obtaining the reward. It was hypothesized, therefore, that this group would attempt fewer examples and project less achievement imagery than any of the other groups. Consequently, further analysis was undertaken comparing the .9, .5, and "prompt" .1 members ( $N = 12$ ) with the "tardy" .1 subjects ( $N = 9$ ) for arithmetic examples undertaken and correctly solved, and their manifestation of achievement imagery. No significant differences were found for either arithmetic examples attempted or solved correctly, though the trend was in the predicted direction. The findings for achievement imagery as shown in Table 4 indicate that the late group projected less achievement imagery ( $M = -.82$ ) than any of the other groups, though only the .5 group showed a significant superiority for the total of 9 cards. Each of the other groups scored significantly more imagery for the low cards, however, and the .5 group manifested significantly more *n*-Ach for the medium set as well. Last, for each of the groups a correlation was computed between *n*-Ach and arithmetic examples attempted. As with the earlier study by Murstein and Collier, the correlations for the .1 ( $r = .00$ ) and .5 groups ( $-.20$ ) proved to be nonsignificant, while that for the .9 group reached significance at the .01 level ( $r = .48$ ).

TABLE 3

MEAN NUMBER OF ARITHMETIC EXAMPLES ATTEMPTED,  
STANDARD DEVIATIONS AND *t* VALUES FOR  
GROUP COMPARISONS

Group	Mean examples attempted	<i>SD</i>	Group comparisons	<i>t</i>
.9	376.4	92.1	.9 vs. .5	.21
.5	373.6	104.0	.5 vs. 1.	1.50
.1	349.6	117.1	.9 vs. .1	1.32

## DISCUSSION

The results of this study support Atkinson's theory with regard to the projection of *n*-Ach, but not for performance on the arithmetic task. This is contrary to the earlier work of Murstein and Collier (1962) and Easter and Murstein (in press) where varying probability groups did not show clear significant differences in achievement imagery. The failure of the arithmetic tasks to differentiate the groups in the present study also runs exactly opposite to the findings of Murstein and Collier (1962) as well as the earlier work of Atkinson (1958b). The explanation of the reversal in the projection of *n*-Ach lies in the present use of cards representing the entire range of achievement expression and thus including a medium-structured subset of achievement-oriented cards. It is these cards which primarily accounted for the differences between the groups. The haphazard selection of cards in earlier studies may well have hit upon cards either too low structured to elicit *n*-Ach, or too highly structured so that both high and low achievement-motivated persons responded to the omnipresent cues. The effectiveness of the low-structured cards in differentiating *all* of the groups ("prompt" .1, .5, .9) from the "late" .1 group further suggests that these cards may be especially effective when dealing with unusually low-motivated persons who reveal themselves by an extremely low rate of manifestation of *n*-Ach. Medium stimulus-pull cards probably elude the largest differences between the groups overall, because they possess sufficient cues to trigger off an achievement response from highly achievement-oriented persons,

TABLE 4

*t* VALUES OF *n*-ACH BETWEEN "PROMPT" .1 GROUP  
.5 GROUP, .9 GROUP AND "LATE" .10 GROUP  
FOR TOTAL SET OF CARDS AND LOW,  
MEDIUM AND HIGH SUBSETS

Group comparisons	Total	Low	Medium	High
"Prompt" .1 vs "Late" .1	.97	1.99*	1.55	.02
.5 vs "Late" .1	2.55**	2.69**	3.95**	.59
.9 vs "Late" .1	1.25	2.23**	1.30	.81

\*  $p < .05$ .

\*\*  $p < .01$ .

but not so much that the low-motivated person is forced to respond out of respect for the reality demands of the picture. The finding that the medium and low cards generally proved more effective than the high-structured cards leads to the question of what is meant by medium stimulus pull for achievement. The view advanced here is that medium pull for achievement refers to the fact that a card elicits a considerable but not preponderant number of themes reflecting *n*-Ach. Such a card is 14 of the TAT which is described by Murray (1943) as "The silhouette of a man (or woman) against a bright window. The rest of the picture is totally black [p. 20]." This picture proved to be the only picture which by itself significantly differentiated the .5 group from the others. It is a picture which elicits themes of achievement with some frequency, but not exclusively so, for themes of escape, depression, and suicide are quite frequent. Another characteristic of a medium stimulus-pull card is that it often clearly depicts *who* is in the picture, but remains ambiguous as to the actions of the characters or *what* is going on. Current work is being done on these questions and will be reported on in future articles.

When attention is turned to the task of explaining the failure of the arithmetic tasks to differentiate the groups, the task becomes more formidable. Possibly, the groups were differentially aroused by the varying probabilities of winning a reward and reflected this differential arousal state in their manifestation of *n*-Ach. When, however, they convened 2 days later, their motivation may have



elapsed as a function of time, or, they may have become bored with the simple nature of the problems and their great number (600).

If the theory is affected by this time delay and/or the fact that a slight difference in complexity existed between the simple arithmetic problems presented by Atkinson and the even simpler ones which were utilized in the present study, then the theory is in need of greater elaboration before it can be more usefully employed in the study of motivation.

Yet another possibility is that a multi-incentive condition as the present one, arousing both achievement motivation and the desire to acquire a small monetary prize, may have served to depress the achievement incentive for persons high in the achievement motive, a possibility supported by Atkinson and Reitman (1958). These authors noted a decline for high *n*-Ach men in attempted arithmetic problems from an achievement-orientation condition to a multi-incentive one in which an additional cash prize was offered. They also, however, showed an increase in arithmetic performance for low *n*-Ach men under these conditions. Since low *n*-Ach men are held to be particularly sensitive to extrinsic rewards such as money (McClelland, 1961), it would then follow that the .9 group should have manifested the best arithmetic performance, but this was not the case. Rather, no significant differences were found between any of the groups on arithmetic performance.

Next, we are confronted with the problem of accounting for the positive correlation for the .90 group between achievement imagery and arithmetic performance ( $r = .48$ ), and the absence of a significant correlation for the other groups. In an earlier study (Murstein & Collier, 1962) we obtained a similar positive correlation, but begged off offering an explanation since the result was so unexpected. The repetition of this earlier finding no longer permits this luxury.

It should be noted that this finding is contrary to what would be expected according to Atkinson's theory. The theory suggests that performance and fantasy should be positively correlated at the point of highest achievement motivation, that is, .5. McClelland

confirms this view (1961), showing his highest positive correlations between *n*-Ach and arithmetic performance for a .5 group of Japanese and Brazilian subjects, whereas the .75 group produced all negative correlations, though these were low.

A possible explanation is that a situation where the probability of success is very great affects achievement imagery and arithmetic performance in the following way. First it imposes a "set" in manner of approach to both tasks which slightly reduces variability. This "set" is not necessarily one to strive for achievement, since the mean score of the .9 group is not high on either task relative to the other groups. Rather, the prospect of success strikes the group in different ways, some individuals presumably being highly motivated for both tasks, whereas others are not greatly motivated in their approach to both tasks. The important fact is that the high probability of success does have considerable impact with regard to both tasks.

Next, we may ask whether the results were influenced by the length of our test. Reitman and Atkinson (1958) and the present author (1963) have indicated that the use of large numbers of cards may serve to mask significant findings that are present when only the first few cards are considered. Reitman and Atkinson, for example, found that *n*-Ach scores to the first four of a series of eight cards predicted differences in performance on arithmetic tasks more successfully than when all eight cards were considered. Whether this effect occurs from a drainage of *n*-Ach on the first four cards or whether, in the desire not to be repetitive, persons expressing achievement themes on the early cards avoid doing so on the later cards, is a moot question. In any event, is it not possible that the lack of correlation between the achievement imagery and arithmetic scores for the .5 group was affected by the fact that we used nine cards instead of four? This possibility seems unlikely when it is recalled that the .9 group showed a significant correlation despite the use of the long series.

We may further ask whether the superiority of the medium set accrued from the



fortunate circumstance that the medium set appeared earlier on the series than the other sets. It will be recalled, however, that the order of cards arrived at stipulated that each subset may be represented only once in any cluster of three cards. The order of cards according to stimulus value was in fact H, M, L, L, M, H, H, M, L. The first four cards are represented by only one medium card and that one was not the most effective within the medium subset. It may be concluded, therefore, that the findings of this study were not seriously affected by the length of the test.

In sum, does the fact that we have found partial confirmation of the theory of risk taking behavior, but also partial lack of confirmation, indicate that the theory is in need of revision? Perhaps so, but before one may justifiably reach this conclusion, it would appear necessary to more rigorously control certain conditions which this experiment clearly indicates may affect the results. These are:

1. The stimulus properties of the cards should be assessed, for it appears that different levels of structure have varying degrees of ability to differentiate those high and low on *n*-Ach.

2. The number and intensity of motives elicited should be controlled. This study, for example, suggests that achievement imagery and arithmetic performance are related to each other most strongly when the probability of monetary gain is quite high, (.9), whereas a moderate possibility of gain (.5) has a strong effect on imagery, but not on performance.

3. Last, the performance task needs to be independently assessed in terms of its arousal

of the achievement motive for a particular sample. McClelland's work (1961) illustrates that arithmetic examples having high arousal properties for one national group may possess little incentive for other nationalities. So too, in the present experiment, it is tenable to believe that the results obtained, which differ from those of Atkinson (1958), are due to a difference in the difficulties of the problems which, accordingly, affected their motivation educing properties.

## REFERENCES

- ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, New Jersey: Van Nostrand, 1958. (a)
- ATKINSON, J. W. Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, New Jersey: Van Nostrand, 1958. Pp. 288-305. (b)
- BIRNEY, R. C. The achievement motive and task performance: a replication. *J. abnorm. soc. Psychol.*, 1958, 56, 133-135.
- EASTER, L. V., & MURSTEIN, B. I. The relationship of expectancy of reward to achievement fantasy on the TAT. *J. consult. Psychol.*, (in press).
- MCCLELLAND, D. C. *The achieving society*. Princeton, New Jersey: Van Nostrand, 1961.
- MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard Univ. Press, 1943.
- MURSTEIN, B. I. *Theories and research in projective techniques: Emphasizing the TAT*. New York: Wiley, (1963).
- MURSTEIN, B. I., & COLLIER, H. The role of the TAT in the measurement of achievement as a function of expectancy. *J. proj. Tech.*, 1962, 26, 96-101.
- REITMAN, W. R., & ATKINSON, J. W. Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, New Jersey: Van Nostrand, 1961. Pp. 664-683.

(Received December 17, 1962)

## EXTRAEXPERIMENTAL EFFECTS IN VERBAL CONDITIONING<sup>1</sup>

WILLIAM R. WIMSATT

AND

NORRIS D. VESTRE

*Minneapolis Rehabilitation Center*

*University of Minnesota*

74 male psychiatric patients, selected according to criteria thought to be related to conditionability, participated in experimental sessions in which they received the verbal reinforcer "good." 1 group was reinforced for responding to MMPI Si items in the scored direction, 1 group was reinforced for responding in the nonscored direction, and a 3rd group received no reinforcement. A significant conditioning effect was found but no extra-experimental or generalization effects were demonstrated either in respect to a correlated self-administered temperament scale or in respect to withdrawal symptoms as rated on a correlated behavior scale. The practicality of employing verbal reinforcement as a means to effect changes in basic personality variables is not supported, nor is its interpretation as an "automatic strengthening" phenomenon sustained.

Several investigators in the area of verbal conditioning have been intrigued by the possibility that this phenomenon, properly understood, might come to serve at least in part as an explanatory principle in accounting for benefits said to result from psychotherapy. Verbal conditioning studies to date, however, as reviewed most lately by Salzinger (1959), have focused primarily on the problem of demonstrating the existence of the phenomenon itself and establishing the conditions under which it will appear; such studies have thus, for the most part, confined themselves to considerations of direct effects observable within the conditioning trials.

It is clear, however, that the proper evaluation of any procedure which is to be called therapeutic must ultimately involve an account of the patient's activity outside of the treatment sessions. Likewise, if verbal conditioning techniques are to be shown analogous to, or argued to be in principle the same as psychotherapeutic techniques, it is essential that behavioral changes occurring outside the conditioning sessions be demonstrated.

While the relatively few studies which have been conducted recently in an effort to clarify the area of extra-experimental effects in verbal conditioning have been perplexing in a num-

ber of ways, there seems to be fair agreement at least in regard to the conditionability of some response categories based on personality dimensions. Oakes and Droge (1960) have reported a successful attempt to condition responses to Social Introversion (Si) scale items on the Minnesota Multiphasic Personality Inventory (MMPI) using the verbal reinforcer "good," while Nuthmann (1957) was able to influence responses to a self-acceptance inventory and Hildum and Brown (1956) were able to manipulate responses to an attitude questionnaire with the same reinforcer. Thus it seems well established that at least some conceptual categories, defined in terms of certain personality variables, can be identified as response classes when appropriately reinforced.

The question is still very much at issue, however, as to whether personality variables themselves, as manifested in other behavior and as measured by other devices than those involved in the conditioning routine, can be altered by verbal reinforcement. That is to say, does being able to influence the score recorded on some self-report scale by means of verbal conditioning techniques imply that the trait upon which the subject is assumed to be reporting is itself influenced? An affirmative answer of sorts is provided by Singer (1961), who was able to condition prodemocratic responses to the California F scale items and was then able to show that scores later obtained on the California E

<sup>1</sup> Data for this study were collected while both authors were at the Veterans Administration Hospital, St. Cloud, Minnesota.

scale were affected. This generalization effect was observed, however, only in the case of that group the members of which completed the *E* scale while the experimenter remained in the room.

A negative conclusion, on the other hand, would seem to follow from an experiment performed by Rogers (1960). While successful in the conditioning of self-reference statements in what he calls a "quasi-therapy" setting, Rogers found the scores on two postsession adjustment scales (Taylor Manifest Anxiety Scale and *Q* Sort Adjustment Test) to be apparently unaffected; likewise, frequency of self-reference statements outside of the experimental sessions was not significantly increased.

Claiming with Singer (1961) that the effects of the reinforcing verbal stimuli are not confined to the conditioning sessions are Ullmann, Krasner and Collins (1961), who reinforced "emotional words" in four experimental sessions and used a behavior rating scale, the Palo Alto Group Therapy Scale, as a dependent measure. Unfortunately, their conclusions are based on an incomplete statistical analysis. A reanalysis of their data along more generally accepted lines by the present authors failed to reveal significant differences in the dependent variable, though the obtained differences were in the "right" direction.

Dulany (1961) raises another issue when he questions the adequacy of operant conditioning principles to describe the data obtained in verbal conditioning studies, and suggests that subjects who condition may actually be operating under self-instructional sets founded on their hypotheses about what they are supposed to be doing. The possibility that verbal conditioning is best conceived as a kind of problem solving or concept formation phenomenon is interesting, but while Dulany concerns himself mostly with the attempts to demonstrate awareness, or the lack of it, on the part of subjects who proved conditionable in various studies, it may be there is better evidence upon which to judge the issue. The confirmation or disconfirmation of the possibility of obtaining extraexperimental effects is important in this

regard, it would seem, since confirmation would support an "automatic strengthening" interpretation of the phenomenon while nonconfirmation would suggest an "hypothesis formation" interpretation.

The experiment here described was designed to investigate possible extra-experimental effects of verbal conditioning techniques applied to hospitalized psychiatric patients. In particular it was intended to test whether the selective reinforcement of responses to items of the *Si* scale of the MMPI can be shown to affect scores on a correlated self-report inventory or to affect withdrawal behavior as reported on a ward rating questionnaire also known to be correlated with *Si*.

## METHOD

### Subjects

The subjects were 74 male patients from the St. Cloud, Minnesota, Veteran's Administration Hospital who were selected to meet the following criteria: (a) 55 years of age or less; (b) known to be literate; (c) not suffering any known impairment of the central nervous system; (d) at least of average intelligence by the most recent available test results; and (e) not receiving unusually large drug dosages. "Unusually large" was defined as a daily dose greater than 600 milligrams of chlorpromazine (Thorazine) or the equivalent in other drugs. The foregoing criteria were applied in order to obtain a group which it was felt would probably be conditionable. The mean age of the sample was 36 years and the mean length of current hospitalization was 5.5 months. With respect to diagnosis, the sample included 48 schizophrenic disorders (65%), 4 other psychotic disorders (5%), 17 neurotic disorders (23%), and 5 personality trait or pattern disturbances (7%).

### Procedure

One hundred  $3 \times 5$  cards were prepared upon 70 of which were typed items scored on the *Si* scale of the MMPI. The remaining 30 were filler items also taken from the MMPI but not scored on any scale; these were included to disguise the character of the items as a whole and to attenuate the pattern of reinforcement for those experimental subjects who happened to have a high initial operant level for responses reinforced in the group to which they were assigned.

The patients were randomly assigned to three groups: Experimental I, Experimental II, and Control. Each subject participated in two experimental sessions spaced 2 days apart. In each session a subject read aloud and responded to 50 of the 100 items. Each set of 50 was composed of 35 *Si* items randomly interspersed with 15 of the filler



items. The cards were presented one at a time by the experimenter and appeared in the same order for every subject. Each response to an *Si* item was recorded. A typical session lasted about 15 minutes.

In the case of those subjects assigned to the Experimental I group, each *Si* item read and followed by a response in the scored direction was reinforced by the experimenter's "good," uttered in as encouraging a manner as possible. No other items, *Si* or filler, were so reinforced. Experimental II was treated just as Experimental I with the exception that only those responses which were *opposite* to the scored direction with respect to the MMPI *Si* scale were reinforced. No responses were reinforced for members of the control group. Thus, Experimental I subjects were reinforced for "introverted" responses, Experimental II for "extroverted" responses, and control subjects received no reinforcement. All subjects were seen by the same experimenter.

Immediately following the second experimental session each subject was required to complete a self-administering survey form which contained all the items scored on the scales *G*, *E*, *S*, and *A* of the Guilford-Zimmerman Temperament Survey (GZTS) (Guilford & Zimmerman, 1949). These scales have been shown to be significantly correlated with *Si*, at least in a normal male population (Miller & Cottle, 1955). Subjects completed this form seated at a table in a room separate from the experimenter's office, where the experimental sessions took place, and in the presence only of other patients similarly occupied.

Preceding the first experimental session and following the second, each subject was rated on the Withdrawal scale of the Psychotic Reaction Profile (PRP) (Lorr, 1961) by a ward attendant. In addition, a follow-up rating 1 week after the second session was obtained in order to provide for the possibility that some passage of time might be required in order to permit the occurrence of instances wherein raters could observe changes of behavior.

## RESULTS

Table 1 shows the mean *Si* scores obtained during the experimental sessions, the mean total scores on the four GZTS scales, and also the mean score on the *S* scale of the

TABLE 1

GROUP MEANS FOR EXPERIMENTALLY OBTAINED *Si* AND FOR POSTEXPERIMENTAL GZTS SCORES

Scales	Experimental I	Control	Experimental II
MMPI <i>Si</i>	( <i>N</i> =74) 29.83	28.50	26.08
GZTS <i>S</i>	( <i>N</i> =65) 17.77	17.83	17.38
GZTS total	( <i>N</i> =65) 69.67	65.67	68.76

TABLE 2

ANALYSIS OF COVARIANCE FOR *Si* SCORES ADJUSTED ON PRE-EXPERIMENTAL *Si* SCORES

Source	<i>df</i>	Adjusted <i>MS</i>	<i>F</i>
Between	2	224.33	3.63*
Within	58	61.88	

\*  $p < .05$ .

GZTS for the two experimental groups and for the control group. The reduced number of cases associated with the GZTS scales is due to the exclusion of nine cases whose scores were invalidated by excessive responding in the "can't say" category.

An analysis of variance performed on all 74 cases for differences between groups with respect to *Si* failed to yield a significant *F* ratio. For 62 members of the original sample, however, pre-experimental *Si* scores were obtained from hospital records for use in covariance analysis. The mean duration between these pre-experimental and the experimental obtained *Si* scores was 3.8 months. Table 2 shows the results of an analysis of this latter data by the covariance method with adjustment made for pre-experimental *Si* differences.

While neither the Control group-Experimental I or the Control group-Experimental II differences were significant, a *t* of 2.68 (Lindquist, 1953) was obtained for the difference between the adjusted means of Experimental group I and Experimental group II, indicating the presence of the conditioning effect.

As is seen, while the three groups were ordered as expected with respect to *Si* scores, and a significant *F* was obtained, the magnitude of the differences between groups was not very large. However, the experimental procedures were such that for subjects having a low operant level initially for responses to be reinforced in the group to which they were assigned, many items in the first session might be read and responded to before the reinforcement could begin to be applied. Reasoning thus, a further analysis was performed involving only the last 15 items (last half of

the second treatment session). For these items, a significant  $F$  was obtained (.01 level) for differences between groups, and a comparison of the Experimental I group with the control group yielded a significant  $t$  (.05 level). While the Experimental II group made fewer  $Si$  responses than the control group in this analysis as in the preceding analysis involving all  $Si$  items, the difference was still not statistically significant.

The covariance analysis for differences between groups in respect to the GZTS data revealed nonsignificant results both for the total score ( $F = .80$ ) for the four GZTS scales administered and for the  $S$  scale ( $F = .57$ ) taken separately. The  $S$  scale was analyzed separately because it was found that for this sample the correlation of  $S$  with pre-experimental  $Si$  was higher ( $-.56$ ) than was the correlation of pre-experimental  $Si$  with the total score for the four scales administered ( $-.49$ ). The correlation between  $S$  scale scores and  $Si$  scores recorded during the experimental sessions was  $-.83$ , the .99 confidence limits about which coefficient being  $-.912$  and  $-.684$ . These correlation coefficients are all different from zero beyond the .01 confidence level, and in addition, the difference between the two correlations of  $S$  and  $Si$  yields a  $CR$  of 2.83,  $p < .01$ .

Similar analyses were performed on the PRP Withdrawal scale scores adjusted on the preexperimental withdrawal scale scores. In respect to this behavior rating scale, no nonchance differences are exhibited either on the first postexperimental rating or on the second.  $F$  ratios of .99 and .36, respectively, were obtained for the two ratings. An earlier study by one of us (NDV) on 78 schizophrenic patients yielded a correlation of .40 between  $Si$  and the PRP Withdrawal scale.

#### DISCUSSION

Since no extraexperimental effects were demonstrated, it would appear that no temperament or personality factor was actually influenced by the verbal conditioning procedure. The elevation or depression of  $Si$  scores by this conditioning technique is then of no more psychological consequence than if such elevation or depression were produced

by a mischievous test scorer, and it seems in effect as if the thermometer has been tampered with while the furnace remains untouched. One interpretation is merely that the generalization gradient is very steep; but perhaps more intuitively appealing is the interpretation that, regardless of the instructions originally given the subject, the verbal conditioning sessions are actually to be considered a type of problem solving task in which the subject learned how to make the experimenter say "good." The failure of such problem solving behavior to persist in other situations is, thus, less surprising.

Since the correlation between GZTS  $S$  scale scores and  $Si$  scores recorded during the experiment was so strong, we were surprised at first to have obtained such a small  $F$  for the  $S$  scale scores while a significant  $F$  was obtained for differences in  $Si$ . The explanation for this apparent paradox is the very steep regression line for  $S$  on  $Si$ .

The significant difference between  $r$ 's for  $S$  scale— $Si$  (pre-experimental) and  $S$  scale— $Si$  (obtained during experiment) seems similar to differences Singer (1961) found between pre- and during-experiment correlations and which he suggested might be explained by supposing his subjects to have learned during the experiment the concept theoretically connecting the  $E$  and  $F$  scales. However, the significant difference between  $r$ 's here obtained might be more simply explained by the fact that, while the second correlation of  $S$  and  $Si$  was computed on measures taken within 3 days of one another, a lapse of about 4 months, for the typical subject, occurred between the registering of the pre-experimental  $Si$  scores and the administration of the GZTS scales. Thus, any changes in respect to these variables occurring within the intervening months would tend to magnify the difference between the correlations.

Finally, there remains, of course, the possibility that the conditioning effect achieved was of such low strength that real extraexperimental concomitants went undetected by reason of their being of such slight magnitude. Perhaps more conditioning trials could have increased the effect to an extent that a generalization effect would have appeared.

REFERENCES

- DULANY, D. E., JR. Hypotheses and habits in verbal "operant conditioning." *J. abnorm. soc. Psychol.*, 1961, 63, 251-263.
- GUILFORD, J. P., & ZIMMERMAN, W. S. *The Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply, 1949.
- HILDUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, 53, 108-111.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- LORR, M. *The Psychotic Reaction Profile Manual*. Beverly Hills, Calif.: Western Psychological Services, 1961.
- MILLER, R. S., & COTTLE, W. C. Evidenced relationships between MMPI and GZTS scales: An adult male sample. *U. Kan. Bull. Educ.*, 1955, 9, 91-94.
- NUTTMANN, ANNE M. Conditioning of a response class on a personality test. *J. abnorm. soc. Psychol.*, 1957, 54, 19-23.
- OAKES, W. F., & DROGE, A. E. Operant conditioning of responses to social introversion scale items on the MMPI. *Psychol. Rep.*, 1960, 6, 223-225.
- ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, 60, 247-252.
- SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, 61, 65-94.
- SINGER, R. D. Verbal conditioning and generalization of prodemocratic responses. *J. abnorm. soc. Psychol.*, 1961, 63, 43-46.
- ULLMAN, L. P., KRASNER, L., & COLLINS, BEVERLY J. Modification of behavior through verbal conditioning: Effects in group therapy. *J. abnorm. soc. Psychol.*, 1961, 62, 128-132.

(Received January 7, 1963)



## COUNSELOR ATTITUDES AND CLIENT BEHAVIOR<sup>1</sup>

IRENE E. WASKOW

*Psychopharmacology Service Center, National Institute of Mental Health*

The hypothesis of this study was: Degree of counselor acceptance, interest, nonjudgmentalness, and expressiveness is directly related to the degree to which the client, fairly early in counseling, both discusses and expresses his feelings. Ratings were made on the basis of interview material. Counselor attitudes were rated in 3 different conditions: reading transcripts, listening to tape recordings, and listening to filtered tapes (with intelligibility eliminated). Counselor attitudes that were reliably rated were related to the client measures by means of partial correlations. Significant ( $p < .05$ ) relationships were found between counselor judgmentalness, in all rating conditions, and client discussion of feelings. These relationships are in a direction opposite to that predicted, and attempts are made to clarify this finding.

This study is concerned with the client's learning, fairly early in a series of counseling sessions, that the counselor accepts him as a person and that he is "safe" to say, to think, and to feel things that have not been permissible in the past. More specifically, the study focuses on the relationship of certain counselor attitudes to aspects of client behavior that might reflect such learning that the counseling situation is "safe." Ratings of both the counselor attitudes and client behavior are based on tape recordings of actual counseling sessions.<sup>2</sup> On the basis of theoretical writings in the areas of counseling and psychotherapy and of several pilot studies, four counselor attitudes were chosen as both relevant to this problem and as reliably rateable: the counselor's acceptance of the client, his interest in the client, his judgmentalness, and his expressiveness. Since it seemed that vocal dimensions might

be used by the client as cues to the counselor's attitudes, these attitudes were rated on the basis of filtered tape recordings (discussed below), as well as unfiltered tapes and transcripts. The two dimensions of client behavior selected as possible signs of the client's achieving some feeling of "safety" in the counseling setting were the extent to which the client deals with (i.e., discusses) his feelings, and the amount of feeling he actually expresses.

Thus, the specific hypotheses of this study are as follows: the degree of counselor acceptance, interest, nonjudgmentalness, and expressiveness will be directly related to the degree to which the client, fairly early in the counseling sessions, both discusses his own feelings and actually expresses his feelings. Predictions for the first three counselor attitudes can be easily made on the basis of a broad range of theoretical writings on counseling and psychotherapy (e.g., Fenichel, 1941; Menninger, 1958; Rogers, 1951, 1957). Although authors of different orientations may use different terminology—e.g., Fenichel's "tolerance" and Rogers' "unconditional positive regard"—they generally agree on the importance of these attitudes. The prediction in regard to the fourth counselor characteristic, expressiveness, is more equivocal. The use of this variable is based partly on the recent work of Butler's group<sup>3</sup>, on the expressiveness of the therapist's speech. Since the present study is

<sup>1</sup> This investigation was conducted at the Laboratory of Psychology, National Institute of Mental Health, during the tenure of a Post-doctoral Research Fellowship from the National Institute of Mental Health. Appreciation for their constant help and encouragement during the course of this project is expressed to David Shakow, Allen Dittmann, and Paul Bergman. The author is also grateful to Thomas Ewing of the University of Illinois Counseling Service and to John Butler of the University of Chicago Counseling Center and their staffs for their cooperation in this study.

<sup>2</sup> Although the terms "client," "counselor," and "counseling sessions" are used throughout this report, it is felt that the nature both of the clients' problems and motivations and of the counseling methods used in the current sample (to be discussed later) renders these interactions similar to those of "patients" and "therapists" in "psychotherapy."

<sup>3</sup> Laura N. Rice, Alice K. Wagstaff, and J. M. Butler, unpublished manuscript, Classification System for Therapist Responses, Univ. Chicago Counseling Center, January, 1960.

of an exploratory nature, any direction of relationship between the counselor attitudes and client behavior variables will, of course, be of interest.

## METHOD

### *Materials Used for Ratings*

The basic material utilized for testing the hypotheses of this study consisted of tape recordings of 28 counseling interviews, two for each of 14 client-counselor pairs. For each pair, the first interview was included, as well as one interview between the twentieth and thirtieth. Also available for most cases was the second interview and an additional interview between the twentieth and thirtieth. This interview material was obtained from two university counseling centers. The counselors included nine males and five females. There were no requirements in the request for tapes regarding amount of experience or other counselor qualifications. The aim was to obtain representative counseling interactions. The counselors were regular employees of these centers and the counselor sample covered a wide range of experience. The orientations of the counselors were varied, including three or four who might be considered "client-centered," but more who were generally eclectic in their procedures. Most of the clients were undergraduate students. They presented a variety of problems, largely neurotic in nature. All of them dealt in their interviews with general problems in adjustment, rather than merely focusing on educational or vocational difficulties. The fact that a minimum of 20 interviews was required by the original specifications in the request for tapes to be used in the study also set a minimal level in terms of the depth or intensity of the counseling experience. There are undoubtedly other effects of this requirement, including a restriction of the client sample to a group of moderately successful cases, since many "failure" cases ordinarily drop out before this point.

### *Rating Counselor Attitudes*

Excerpts of counselor speech were selected from either the first or second interview of each of the counseling cases. The first interview (or second, in the five cases for which there was poor audibility or too few usable segments in the first interview) was used in order to eliminate, as far as possible, the effect of the client on the counselor's attitudes. The assumption was made, however, that these attitudes and characteristics would be fairly constant throughout the early interviews. Four 3-response units were selected for each counselor, and these were transferred to new tapes with all client sounds omitted. The four 3-response units selected for each counselor were the first three responses after 5 minutes, 15 minutes, 25 minutes, and 35 minutes that met or could be cut to meet the following requirements: (a) no response shorter than five words or longer than 45 seconds, (b) no 3-response unit longer than 1.5 minutes, (c) no 3-response unit shorter than 15 seconds.

These requirements assured that the rating task would not be too long (1.5 to 2 hours) and that there was

enough material (as indicated in pilot work) on which to base each rating. The 56 segments (4 for each of 14 counselors) were arranged in a random order with the restrictions that there be no two consecutive responses of the same counselor and that each counselor be represented once in each 14 consecutive segments.

Ratings were made in three different experimental conditions: (a) listening to tape recordings of the counselor responses, (b) listening to filtered tape recordings, and (c) reading transcripts. Before starting to rate, and in order to establish a frame of reference, the raters heard (or read) the first 14 segments, which were then reheard (or reread) as the rating task began. For the filtered condition, Soskin's technique (Soskin & Kauffman, 1960) was utilized and frequencies above 450 cps were either eliminated or greatly attenuated. The speech was thus rendered unintelligible, while most voice quality characteristics were retained. This technique had already proved its usefulness in several studies of content-free speech (Starkweather, 1956a, 1956b). In the present study the ratings based on filtered tapes could be taken as measuring only that which is conveyed by the vocal characteristics of the counselor's speech. To assure the use of verbal as well as vocal cues in the unfiltered tape condition, raters had transcripts before them as they listened to the tape.

Before beginning the study, the raters studied the definitions of the characteristics to be rated and a description of the rating system to be used. The definitions, the same for raters in all three conditions, are reproduced below. The rating system had been found in pilot studies to result in fairly normal distributions of ratings.

## DEFINITIONS OF COUNSELOR ATTITUDES

<i>Accepting</i>	<i>Rejecting</i>
Accepts the client as a person; has respect for and values him; recognizes his worth as an individual, regardless of limitations.	Is critical of the client as a person; treats him as inferior, to be disregarded or manipulated; is condescending.
<i>Expressive</i>	<i>Nonexpressive</i>
(As popularly defined) has life and color; is variable; does not seem dull.	"Blah"; seems dull, lifeless, little variability.
<i>Judgmental</i>	<i>Nonjudgmental</i>
Seems evaluative; sets self up as the judge rather than client as judge; seems to push own views as correct ones; acts as if he knows the answers.	Seems neutral, nonjudging, nonevaluative; acts as if client is the final judge.
<i>Interested</i>	<i>Indifferent</i>
(As popularly defined) concerned; alert, attending to and interested in the client's communications.	Apathetic, bored; not particularly concerned or involved in what's going on.



## SAMPLE RATING SCALE AND INSTRUCTIONS

Rejecting			Accepting
1	2	3	
Nonexpressive			Expressive
1	2	3	
Nonjudgmental			Judgmental
1	2	3	
Interested			Indifferent
1	2	3	

In making a rating, please circle the number that corresponds to your judgment, e.g., if you feel the counselor seems accepting, circle (3), rejecting (1), or some mid-position between accepting and rejecting (2). If the responses do not clearly belong in one of these categories, an arrow may be drawn to another category, e.g., a response that is somewhat, but not clearly, accepting may be rated 2←(3). Please try not to pile up your ratings at any one point, but rather try to spread them through the range of 7 possible ratings: (1); (1)→2; 1←(2); (2); (2)←3; 2→(3); (3)

The four characteristics were rated in different random orders for different 3-response units. After each segment of 3 responses was read or heard, the rater was asked to make a global rating for the 3 responses on each of the 4 attitude dimensions. He was given as much time as he found necessary. The instructions included a request that the rater "try to make the ratings as if you were the client to whom the counselor is speaking."

The raters were college students and recent college graduates (research assistants and normal control subjects in National Institutes of Health projects), without much sophistication about counseling or psychotherapy. It was hoped that their responses would be somewhat similar to those of actual counseling center clients. There were six raters in each of the three rating conditions, and these groups of raters were adequately matched for mean age and number of years of college. The six groups of ratings for each condition were obtained in two sessions, with three raters run simultaneously.

All analyses were based on the averages of the six raters' mean ratings (i.e., mean of 4 separate ratings) for each counselor. Reliabilities of these average ratings (Ebel, 1951) were at a level acceptable for exploratory work for a number of the characteristics: Filtered: judgmentalness,  $r = .76$ , expressiveness,  $r = .64$ ; Trans: judgmentalness,  $r = .81$ , acceptance,  $r = .70$ , interest,  $r = .66$ ; Unfiltered: judgmentalness,  $r = .64$ , interest,  $r = .62$ , expressiveness,  $r = .61$ . Only these three groups of ratings were then related to the measures of client behavior.

The average ratings of the different characteristics were also intercorrelated. Strong relationships ( $r > .60$ ) were found between interest and expressiveness in all three rating conditions, between expressiveness and judgmentalness in the filtered condition, and between acceptance and interest and between acceptance and

judgmentalness (the last, a negative relationship) in the transcript condition. Although these correlations probably reflect actual relationships among the four characteristics rated, they undoubtedly are spuriously high because all four characteristics were rated at the same time. Intercorrelations among ratings in the three conditions for each of the characteristics revealed only three strong relationships: for judgmentalness, between filtered and unfiltered ratings, and between unfiltered and transcript (there was also some relationship— $r = .49$ —between unfiltered and transcript ratings for judgmentalness) and for expressiveness, between unfiltered and filtered ratings.

## Rating Client Behavior

Segments of client behavior were selected from both the first interview and some interview between the twentieth and thirtieth for each client. Segments of the first session were rated in order to get a base level on both client variables, before one could expect any changes to occur in response to the counselors' attitudes. It was felt that such changes might be discernible at some point between the twentieth and thirtieth interviews; for this reason the tapes obtained from counseling centers had included the first session and one between the twentieth and thirtieth. If two interviews during this later period were available, the one closer to the twenty-fifth was used.

From each tape used, three 5-minute segments of client speech were extracted, one each from the first 15 minutes, second 15 minutes, and third 15 minutes of the tape (always starting as close to the beginning of the 15-minute period as possible). All counselor responses (except an occasional "Um-hmm" or one word or overlap of two or three words that were almost impossible to omit) were removed, as was the first sentence of any client communication that was simply an answer to a direct counselor question. All pauses between client and counselor communications were omitted, as were all silences, after the first 5 seconds, within the client speech. For both rating methods, a rating was made after each 15 seconds of client speech, thus totalling 60 ratings for each 15-minute segment. A separate tape was made for each group of three 5-minute segments of client speech and these were rated in a random order.

For ratings of the extent to which the client is "dealing with his own feelings," a tape of three 5-minute segments of a client's speech was heard continuously, and a score was made after each 15-second segment signifying whether the client was dealing mainly with his own feelings (✓), not dealing mainly with his own feelings (—), or equally dealing and not dealing with his own feelings (?). After the entire 15-minute tape was heard, it was replayed, and ratings were changed if necessary; this second group of ratings comprised the actual data used. Since one-half credit was given for the (?) category, the total scores could range from 1 to 60, with half-step intervals. After some preliminary practice with the scale, a detailed list of instructions had been developed, including, e.g.:

"Dealing with feelings" means any type of dealing with them; that is, they may be merely discussed rather than expressed in tone of voice, etc.; this still counts as



dealing with feelings; what is important is that the client's own feelings are the subject matter under consideration. As can be seen, the scores on this rating system are essentially a measure of the amount of time during which the client primarily discusses his own feelings.

Again, in making ratings of the "amount of feeling expressed," 15 minutes of client speech was heard continuously and ratings were made each 15 seconds. Here, ratings were based on a 3-category system: N for no feeling expressed, M for a moderate amount of feeling expressed, and I for an intense amount of feeling. Arrows could also be drawn from one category to the next, in a system similar to that used for rating counselor attitudes. Here, though, the ratings were converted to a 5-point scale (which, in pilot work, seemed the most reliable method), with  $\textcircled{N} = 1$ ;  $\textcircled{N} \rightarrow \textcircled{M}$  and  $\textcircled{N} \leftarrow \textcircled{M} = 2$ ;  $\textcircled{M} = 3$ ;  $\textcircled{M} \rightarrow \textcircled{I}$  and  $\textcircled{M} \leftarrow \textcircled{I} = 4$ ; and  $\textcircled{I} = 5$ . The 60 ratings for each client were then averaged. A short list of instructions was developed for use with this rating system, with major emphasis on ignoring content and concentrating on the amount of feeling actually being expressed.

All 28 tapes (two for each of 14 clients) were rated by the author, according to the systems just outlined, and these ratings comprised the basic data on client behavior. To assess interrater reliability, a second rater rated 14 tapes (which included one of each client and seven early and seven late sessions, but which were otherwise randomly selected) on the first client variable, discussion of feelings ( $r = .93$ ) and the remaining 14 tapes on the second client variable, expression of feelings ( $r = .67$ ). It should be mentioned that the .67 reliability figure, although not as high as could be desired, is actually higher than one might expect from the narrow range of scores obtained by both raters.

Correlations among the ratings of client material revealed a high relationship between early (first interview) and late (twentieth-thirtieth interview) ratings for each variable ( $r = .74$  between early and late ratings of discussion of feelings and  $r = .68$  between early and late ratings of expression of feelings) and negligible relationships between the ratings of the two variables ( $r = .26$  for the early interview and  $r = .06$  for the later interview).

The procedure followed thus provided us with eight sets of reliable ratings of counselor characteristics in the first or second interview which could be related to ratings of two aspects of client behavior in the first interview and in an interview between the twentieth and thirtieth.

## RESULTS

Sixteen partial correlations were computed, relating each of the eight counselor variables that had been reliably rated with each of the two client variables. The use of partial correlations allowed an analysis of the relationship between counselor attitude ratings and client behavior in the later interview, while controlling for the effect of the initial level of client

TABLE 1  
PARTIAL CORRELATIONS BETWEEN COUNSELOR  
ATTITUDE AND CLIENT BEHAVIOR  
MEASURES

Counselor attitudes	Client behavior	
	Discussion of feelings	Expression of feelings
Filtered:		
Judgmentalness	.57*	-.17
Expressiveness	.15	.01
Unfiltered:		
Judgmentalness	.74**	.06
Expressiveness	-.03	.25
Interest	-.20	.10
Transcript:		
Judgmentalness	.59*	.17
Interest	.02	.51
Acceptance	-.58*	.19

\*  $p < .05$ .

\*\*  $p < .01$ .

behavior. The final results are presented in Table 1.

Since the only consistently significant results were found for the judgmentalness ratings, the means and ranges of these ratings are presented in Table 2. The mean ratings of this variable are approximately the same for the two groups of counselors (representing different counseling centers), although the distributions are somewhat different, with more extreme ratings given to the Illinois counselors, especially in the unfiltered condition.

In interpreting the results, several factors must be kept in mind: A number of relationships are being investigated and one might expect at least one relationship to appear in the significance range by chance; the counselor attitude dimensions are not completely independent and two related attitudes may well have a similar relationship to a third variable; the reliabilities of each of the counselor attitude measures and of the two client measures set upper limits on the degree of their relationship to each other.

Although four relationships are significant, one must note that only one of four counselor characteristics is consistently related to an aspect of client behavior. There is obviously a need for replication of this finding in future studies. One is struck, however, by this consistency of findings in regard to counselor judgment.

TABLE 2

MEANS AND RANGES OF JUDGMENTALNESS RATINGS  
IN THE THREE RATING CONDITIONS

Rating condition	<i>M</i>	Range
Filtered	3.9	3.0-5.6
Unfiltered	4.0	2.9-5.5
Transcript	3.8	1.8-5.3

mentalness. As rated in all three conditions, this variable is significantly related to the amount of client discussion of feelings. The only other significant finding is the negative relationship between acceptance as rated in the transcript condition and client discussion of feelings. Since ratings of acceptance have a high negative correlation with ratings of judgmentalness in the transcript condition, the relationship of acceptance to discussion of feelings will not be dealt with as a separate finding of this study.

There are no significant findings for the second client variable, expression of feelings. Only one of the correlations involving this variable (relating it to counselor interest in the transcript condition) approaches significance. It is difficult to interpret the absence of significant findings for this variable since its reliability ( $r = .67$ ), in conjunction with the never very high reliabilities of the counselor characteristics, sets a fairly low upper limit to the correlations that could be obtained. It is hoped that in future research a more reliable method of rating client expression of feelings can be developed, thus more definitively revealing the relationship or lack of relationship between it and counselor attitude variables.

The absence of significant relationships between four of the counselor attitude measures and client discussion of feelings will similarly not be dealt with here. The fact that judgmentalness was, in all three rating conditions, the most reliably rated characteristic again made more possible its significant relationship with other variables.

In order to rule out the possibility that the number of words spoken by the counselor might be responsible for the major results obtained, double partial correlations were computed for assessing the relationship between mean judgmentalness ratings for each of the 14 counselors

(as rated in each of the three conditions) and client (late interview) discussion of feelings, while controlling both for the initial level of client behavior and for the total number of words spoken by the counselor. The resulting correlations differ very little from the original ones; they are: filtered,  $r = .55$ ; unfiltered,  $r = .76$ ; transcript,  $r = .57$ . Thus, number of words is not a factor that need be considered in interpreting the findings of this study.

## DISCUSSION

The direction of the relationship of counselor judgmentalness to the amount of client discussion of feelings is, in all cases, opposite to the direction predicted, i.e., the amount of client discussion of feelings is greater when the counselor is rated as being more judgmental. This result is somewhat puzzling: regardless of orientation, many theorists would predict that a judgmental counselor would inhibit, rather than facilitate, the client's discussion of feelings.

One question that arises is whether some other counselor characteristic may be highly related both to ratings of judgmentalness and to the client's increasing discussion of his feelings, and whether such a variable may be indirectly responsible for the results found. Two specific possibilities are: (a) that the more judgmental counselor is directly reinforcing the client's talking about his feelings and (b) that the more judgmental counselor is also seen as being more "forceful," and this factor may be responsible for the client's increased talking about his feelings. Both of these possibilities were briefly investigated, in an effort to get a better understanding of the present findings and in order to develop hypotheses for further research.

A system was devised for rating the degree to which the counselor focuses on the client's feelings, with ratings based on the transcript material. Each response in a 3-response counselor unit was categorized as to whether the counselor was focusing at all on the client's feelings (F), not focusing on feelings (N), or whether this could not be determined (?). The three separate ratings for each unit were then systematically combined into one rating on a 7-point scale (e.g., one N and two ? = 3, three F = 7). The 56 counselor units were



rated in the same random order as had been previously used and the four ratings thus obtained for each counselor were averaged. The estimate of interrater reliability for these 14 mean ratings was  $r = .91$ . The relationships of these ratings to transcript ratings of counselor judgmentalness ( $r = .02$ ) and to the degree to which the client discusses his feelings (partial  $r = -.22$ ) were, however, not significant. The degree to which the counselor focused on the client's feelings seemed in no way to influence the relationship between the ratings of counselor judgmentalness and the client's discussion of feelings (a double-partial correlation between counselor judgmentalness and client discussion of feelings in the late interview, controlling for amount of counselor focusing on feelings as well as amount of client early discussion of feelings, yielded an  $r$  of .61). Thus, the present results cannot be interpreted as due to a direct reinforcement of the client's discussion of feelings by the more judgmental counselors.

The other question considered was whether a more forceful counselor might somehow facilitate a client's discussion of feelings and whether such forcefulness might be significantly related to judgmentalness as rated in this study. In a recent study of the factorial structure of therapist responses (Howe & Pope, 1961), one of the factors that emerged was labeled "precision-potency." The five semantic-differential scales with highest loadings on this factor were defined at their end points by the adjectives colorless-colorful, still-vibrant, vague-precise (positive loadings) and energetic-inert and strong-weak (negative loadings). Since it seemed that these scales (with perhaps the exception of vague-precise) might be useful in trying to tap the general "forcefulness" of a counselor's responses, they were applied to the tape-recorded (unfiltered) material.

Four raters made a global rating for each of the 56 3-response units on each of the five semantic differential scales. The general instructions for use of the semantic differential scales were similar to those used in the Howe and Pope study. The reliabilities of the average ratings of the four raters for the 14 counselors were: strong-weak,  $r = .67$ ; colorful-colorless,  $r = .70$ ; energetic-inert,  $r = .88$ ; vibrant-still,

$r = .68$ ; precise-vague,  $r = .84$ ; and for a summed combination of the first four of these scales,  $r = .81$ . The interrelationships among these four scales were very high, ranging from  $r$ 's of .83 to .97, while the relationships of each of the four with precise-vague ranged from  $r$ 's of .46 to .57.

Although each of the potency factors and the combined scale significantly differentiated among the 14 counselors ( $F$  ratios for all at probabilities  $< .01$ ), none of them was significantly related to counselor judgmentalness (unfiltered tape condition) or to client discussion of feelings. In addition, none of them had a significant effect on the relationship between counselor judgmentalness and client discussion of feelings (as estimated by double-partial correlations, as described in the section on counselor focusing on feelings). Thus the relationship found between counselor judgmentalness and client discussion of feelings cannot be attributed to the judged forcefulness (as defined by these "potency" variables) of the counselor.

There was no support for either of the proposed explanations of the judgmentalness results in these two supplementary investigations. We have thus far, however, looked only at the possibility that the judgmentalness results might have been spuriously produced by the presence of other relevant variables. One must also consider the possibility that judgmentalness, as defined, was the significant variable, but that the initial hypothesis was incorrect. It is conceivable that the judgmental counselor may, by the very act of expressing judgments, arouse feelings in the client. If this is the case, however, it is surprising that there was not even a tendency toward a significant relationship between judgmentalness and client *expression* of feeling. Another point arguing against this interpretation is the fact that most of the feelings that the clients *discuss* are not directly related to the counselor or the counseling situation.

In an attempt to find a more defensible interpretation of the results, it might prove illuminating to look again at the significant variable itself, at the specific definition of judgmentalness that was used. A judgmental counselor is defined as one who "seems evaluative; sets self up as the judge rather than client



as judge; seems to push own views as correct ones; acts as if *he* knows the answers." Many theorists in the area of counseling and psychotherapy would hold that an evaluative, judging attitude on the part of the counselor or therapist would hinder the establishing of a safe atmosphere in which the client can freely discuss his feelings. The present definition of judgmentalness, however, includes not only this dimension of evaluation and judgment on the part of the counselor but also an element of assurance that he knows what he is talking about, that he is the authority in this situation. And it is this latter component of the present definition that may be of central importance; it may be that the counselor who acts in this assured fashion provides some security for the client—that it is easier for the client to discuss his feelings with someone who acts as if he will be able to deal with them and may, indeed, offer some solutions. Thus the judgmental counselor, as seen in this light, may be one who, by his own display of assurance, encourages the client to discuss his feelings. It would seem important, in future research, to attempt to separate out the different components of judgmentalness, as the term is used here, and to see if it is indeed this "assurance" component which is significantly related to the client's increased discussion of his feelings.

Although the assurance of the counselor may encourage the client to discuss his feelings, it may not provide him with the freedom actually to experience these feelings in the counseling sessions. One might note here that there was not even a suggestion in the statistical analyses of a relationship between judgmentalness and the client's *expression* of feelings.

The relationship of judgmentalness to *discussion* of feelings may be seen as important in itself if one assumes that for meaningful relearning about feelings to occur, one has to start talking about these feelings. It is important, however, to consider the possibility that discussion of feelings without simultaneous experiencing of them may not be of much therapeutic value. Of relevance here is a study (Gendlin, Jenney, & Shlien, 1960) indicating that the client-centered counselor's judgment of success in therapy is significantly related to his rating of the degree to which

the client does "*express* his feelings," rather than "*talk about* them."

The major purpose of the present research was to relate measures of counselor and client in-therapy behavior, without relating them to outcome measures. Since global success ratings (made by the counselors) were available, however, these were examined in regard to their relationship to the variables of counselor judgmentalness and client discussion of feelings. The trend was in the direction of a negative relationship of outcome with discussion of feelings and (for the Illinois group) with counselor judgmentalness. At any rate, there was clearly no support for a positive relationship between discussion of feelings and counselor's judgment of degree of therapeutic success. These results suggest, then, that although a judgmental counselor can elicit talk about feelings in early counseling sessions, he may not facilitate the patient's dealing with these feelings in a meaningful and therapeutic way. Thus, in terms of the initial formulation of this study, the client may learn to talk about his feelings with a judgmental counselor, though he may not learn from him that the counseling situation is safe for the further experiencing and expressing of feelings that are necessary for therapeutic gain to occur.

Two other aspects of the present results deserve brief mention. The significant relationship with discussion of feelings was found for all three methods of rating judgmentalness. The fact that the highest relationship was found for the unfiltered tape condition ( $r = .74$ ) and somewhat lower and approximately equal relationships for the other two conditions ( $r$ 's of .57 and .59) is especially interesting in light of the high correlations between ratings of judgmentalness in the filtered and unfiltered conditions ( $r = .73$ ) and between ratings in the transcript and unfiltered conditions ( $r = .74$ ), with a somewhat lower correlation between filtered and transcript ratings ( $r = .49$ ). This combination of findings leads to the speculation that each of the dimensions, i.e., verbal (transcript) and vocal (filtered tape), contributes to the perception of judgmentalness, and each makes a contribution which is important in the relationship of this variable to client discussion of feelings.

Also of note is the finding that ratings based on contentless voice quality can relate significantly to other variables in the counseling situation. Previous research (Starkweather, 1956a, 1956b) has utilized this filtering technique in quite different situations.

#### REFERENCES

- EBEL, R. L. Estimations of the reliability of ratings. *Psychometrika*, 1951, 16, 407-424.
- FENICHEL, O. *Problems of psychoanalytic technique*. Albany, New York: Psychiatric Quarterly, Inc., 1941.
- GENDLIN, E. T., JENNEY, R. H., & SHLIEN, J. M. Counselor ratings of process and outcome in client-centered therapy. *J. clin. Psychol.*, 1960, 16, 210-213.
- HOWE, E. S., & POPE, B. The dimensionality of ratings of therapist verbal responses. *J. consult. Psychol.*, 1961, 25, 296-303.
- MENNINGER, K. *Theory of psychoanalytic technique*. New York: Basic Books, Inc., 1958.
- ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
- ROGERS, C. R. The necessary and sufficient conditions of therapeutic personality change. *J. consult. Psychol.*, 1957, 21, 95-103.
- SOSKIN, W. F., & KAUFFMAN, P. E. Judgment of emotion in word-free samples. *J. Commun.*, 1961, 11, 73-80.
- STARKWEATHER, J. A. The communication value of content-free speech. *Amer. J. Psychol.*, 1956, 69, 121-123. (a)
- STARKWEATHER, J. A. Content-free speech as a source of information about the speaker. *J. abnorm. soc. Psychol.*, 1956, 52, 394-402. (b)

(Received January 17, 1963)

## THE CONSTRUCT VALIDITY OF THE PROGRESSIVE MATRICES AS A MEASURE OF SUPEREGO STRENGTH IN JUVENILE DELINQUENTS

WILFRED H. HIGASHIMACHI<sup>1</sup>

*Utah State Department of Health, Division of Mental Health*

Raven's Progressive Matrices (1938 & 1947) scales were used to determine whether they were sensitive in reflecting differences in superego strength and functioning. It was postulated that a high score on the Matrices is associated with a measure of strong superego on the Picture Arrangement Test (PAT), and a low score on the Matrices with a measure of weak superego on the PAT when general ability is held constant. 12 delinquent boys composed the strong superego group, and 12 composed the weak superego group. The combined Matrices scores revealed that the strong superego group had a significantly higher score beyond the .01 level of confidence than the weak superego group. The results give support to the fact that the Matrices are sensitive in reflecting differences in superego strength and functioning in a population of delinquent boys.

The Progressive Matrices "was designed to cover the whole range of intellectual development from infancy to maturity" (Raven, 1951). In England, it ranks second only to Binet's as a test of general intelligence (Slater, 1948). Raven's Progressive Matrices purports to be a quantitative measure of education, and is described as an untimed, nonverbal, nonperformance test of intellectual functioning based on Spearman's *g* factor of intelligence. However, Raven (1954) states that one must not mistake the Progressive Matrices as a test of "general intelligence" but should consider it more as a useful test of "observation or clear thinking."

Since the introduction of the Progressive Matrices to the United States, attempts have been made by others to correlate it with various well-known intelligence tests, such as the Stanford-Binet (Stacey & Carleton, 1955; Stacey & Gill, 1955), Wechsler-Bellevue (Desai, 1955; Levine & Iscoe, 1954), Wechsler Adult Intelligence Scale (Hall, 1957), Wechsler Intelligence Scale for Children (Martin & Wiechers, 1954; Stacey & Carleton, 1955), ACE and Otis (Bolin, 1955), with results having a generally wide range of variability with respective IQ scores.

Bromley (1953) found that the Progressive Matrices failed adequately to account for errors made in disordered thinking of psychiatric cases when the test is regarded as a test of "abstract" intelligence. Watson (1955) noted personality problems affecting performance on the Matrices, and felt that boys with personality problems would perform better if the test were administered individually with adequate encouragement rather than in a group testing situation. Halstead (1943) observed that neurotic subjects who scored a high grade on the Matrices test tended to be less disturbed by their neuroses.

Johnson (1953a, 1953b, 1953c) believed that an objective Raven score may provide some impression of superego capacity, and attempted to investigate the sensitivity of the Progressive Matrices as a clinical predictor of play therapy progress. Johnson (1953a) considers that:

the capacity tapped by this test appears to be the capacity to apprehend the governing principle, to apply it to the untried situation, and to concur in the validity of the solution thus deduced. Readiness to assume such behavior seems commonly related to "character," as well as to level of intellectual potential.

She feels that if the subject is rarely aware of his mistakes, the only pressure upon him is his willingness to become absorbed in solving the problems according to his recogni-

<sup>1</sup> Formerly with the Colorado State Department of Public Health, Community Mental Health Services Division.



tion of the principles of progression. However, Johnson's investigations have not clearly demonstrated that the test performance of her subjects was equated to superego capacity.

There is still much doubt concerning the range of applicability of Raven's Progressive Matrices test. Research concurs that the Matrices cover a wide range of intellectual activities which cannot be comfortably accommodated by Spearman's theory of "general intelligence." A major question is: "What functions of an individual's performance does this test measure?" If personality factors play an important role in a subject's test performance, "Can the Progressive Matrices reflect major differences in personality factors of individuals, such as those of juvenile delinquents, consistently and reliably?" Disorders of superego development have been strongly emphasized in psychoanalytic interpretations of juvenile delinquent behavior (Aichorn, 1951; Eissler, 1949; Forsche, 1956; Friedlander, 1947; Hora, 1952; Redl & Wineman, 1957).

Therefore, this study will attempt to clarify Johnson's belief that the Progressive Matrices test is sensitive as an instrument in providing some impression of superego capacity. It is felt that a person who is able to defer immediate pleasure and become absorbed with a difficult or unpleasant task may possess a strong superego; while a person who is guided by the pleasure principle (in the Freudian sense) and who regards all tasks as unpleasant in which there is no immediate reward may possess a weak superego (Oberndorf, 1951). The purpose of this study was to investigate one particular aspect of the juvenile delinquent's personality, i.e., superego strength, in relation to his performance on the Progressive Matrices.

The selection of subjects and interpretation of performance on the Progressive Matrices in this study were guided by the conception of a strong, restrictive, back-to-work type of superego as measured by the Tomkins-Horn Picture Arrangement Test (PAT). Such a conception of a strong superego was contrasted with a weak, permissive superego which allows liberal, antisocial impulse expression within a work situation.

The PAT (Tomkins & Miner, 1957) consists of 25 plates (each containing line drawings of the "hero" depicted in three different but related situations), with 14 of the plates primarily concerned with the work situation. There are 655 scoring keys, and the logic of the test interpretation rests upon the plate arrangements that are rare or improbable (occurring at the 5% level of probability or less) in any education, intelligence, or age subgroup of the normal sample.

On the basis that the PAT can differentiate juvenile delinquents with strong and weak superego strengths, the major hypothesis of this study may be stated as follows: The Progressive Matrices test is sensitive in reflecting differences in superego strength and functioning; a high score on the Matrices is associated with a measure of strong superego on the PAT, and a low score on the Matrices is associated with a measure of weak superego on the PAT when general ability is held constant.

The specific hypotheses to be tested are as follows:

1. Delinquents who have more superego strength and control over impulse expression will obtain higher scores on the Progressive Matrices (Colored 1947 scale and 1938 scale, separately and combined), as compared to delinquents who have less superego strength and control over impulse expression.

2. Delinquents who have more superego strength and control over impulse expression will be more ego involved with their performance on the combined Progressive Matrices scales, and thus will take a longer time in their performance than delinquents who have less superego strength and control over impulse expression.

3. There will be a greater variability of scores on the combined Progressive Matrices scales among delinquents who have less superego strength and control over impulse expression, than among delinquents who have more superego strength and control over impulse expression.

#### METHOD

The 24 subjects for this study were all selected from a delinquent population of boys from different

racial backgrounds who had come in contact with the law, and were committed by the courts to the Lookout Mountain School for Boys at Golden, Colorado. These subjects were within the age range of 12 years through 16 years possessing an average IQ score (90-110) on the Henmon-Nelson Tests of Mental Ability.

In addition to the above procedures, the following criteria were used in the initial selection of all subjects: (a) At least a sixth grade education, (b) the ability to read and write, (c) a minimum residence of three months at the institution, (d) no physical or organic handicaps, (e) no psychosis, and (f) not receiving any psychotherapy.

By following these procedures with the assistance of the Boys' Counselor at the institution, 48 boys were initially secured. The PAT was administered in a group setting as a screening test. As a safeguard, those subjects who were one response from scoring at the 5% level of probability on the "Strong Superego—Back to Work After Injury, Passivity, Fatigue, Sociophilia, and/or Anger" personality trait were omitted from the "Weak Superego" Group. This resulted in the loss of 14 subjects from the Weak Superego Group. Furthermore, eight subjects were lost through the validity scales of the PAT, and two were lost through illness. The final sample consisted of two groups with 12 subjects in each group:

**Group I (Strong Superego):** Those subjects scoring significantly at the 5% level of probability or less on the "Strong Superego—Back to Work After Injury, Passivity, Fatigue, Sociophilia, and/or Anger" personality trait on the PAT. (This Strong Superego personality trait is defined quantitatively on the PAT as being at the very extreme end of a continuum.)

**Group II (Weak Superego):** Those subjects who did not score significantly at the 5% level of probability or less on the similar personality trait above.

The Strong Superego Group had a mean age of 14.83 years ( $SD = 1.07$ ), mean educational grade of 8.50 ( $SD = 1.12$ ), and mean IQ of 100.17 ( $SD = 6.70$ ). The Weak Superego Group had a mean age of 14.83 years ( $SD = 1.07$ ), mean educational grade of 8.58 ( $SD = 1.24$ ), and mean IQ of 98.83 ( $SD = 5.80$ ).

The subjects who were finally selected for the two groups were all seen individually by the examiner in the institution. Each subject was given the instructions according to Raven's Progressive Matrices' manual for individual test administration, with the examiner emphasizing: "You can have as much time as you like."

After each subject had understood the first pattern in Set A of the Colored Children's Matrices (1947) scale, he was instructed to carry on through the rest of the 71 patterns at his own pace, calling out the answers to the examiner, who scored them on the answer form. Each subject was then timed for the length of his test performance without his being aware of it. The test performance time began when

the subject started with the second pattern in Set A, working through Sets A, Ab, B of the Colored Matrices and continuing through Sets C, D, E of the Matrices (1938) scale until he finished the last pattern in Set E. The subjects were encouraged to give a response to all the 72 matrix patterns.

The Progressive Matrices answer forms were all manually scored by the examiner. As a cross validation study, social histories and adjustment records were evaluated in hopes of finding any pertinent trend which may further differentiate the Strong Superego and Weak Superego groups.

## RESULTS

In analyzing the results of this study, major consideration was placed on comparing the performance on the combined Progressive Matrices (1938 and 1947) scales of delinquent boys possessing a Strong Superego and those possessing a Weak Superego as measured by the PAT.

Statistical analysis of the combined Progressive Matrices (1938 and 1947) scores suggested the use of the Mann-Whitney non-parametric test, since the two groups of scores were obtained from a very select population. The combined Matrices scores revealed that the Strong Superego Group had a significantly higher score beyond the .01 level of confidence than the Weak Superego Group (Group I mean score = 53.92, range = 64-46,  $SD = 5.68$ ; Group II mean score = 38.75, range = 55-16,  $SD = 12.09$ ). The Strong Superego Group also obtained significantly higher scores than the Weak Superego Group on the separate Matrices (1938) scale beyond the .01 level of confidence (Group I mean score = 42.67, range = 52-36,  $SD = 5.44$ ; Group II mean score = 29.92, range = 44-11,  $SD = 10.49$ ), as well as the separate Colored Matrices (1947) scale beyond the .02 level of confidence (Group I mean score = 31.75, range = 35-27,  $SD = 2.45$ ; Group II mean score = 25.92, range = 34-14,  $SD = 5.94$ ).

A comparison of the test performance times of the two groups by the Mann-Whitney Test revealed that the Strong Superego Group took a considerably longer time, and apparently seemed to be more ego involved with their test performance beyond the .001 level of confidence than the Weak Superego Group (Group I mean time = 26.10 minutes, range



= 38 minutes 14 seconds to 19 minutes 54 seconds,  $SD = 5.78$ ; Group II mean time = 15.46 minutes, range = 21 minutes 40 seconds to 9 minutes 38 seconds,  $SD = 3.80$ ).

On the suggestion of McNemar<sup>2</sup> (1957), an  $F$  test was used to estimate the variance of the two groups. It was found that the Weak Superego Group had a greater variability of scores on the combined Progressive Matrices, Sets A, Ab, B, C, D, E, as compared to the Strong Superego Group near the .05 level of confidence using a conservative estimate. Although the separate Matrices (1938) scale, Sets A, B, C, D, E, revealed no significant differences (using a conservative estimate) between the two groups in variability of scores, the separate Colored Matrices (1947) scale, Sets A, Ab, B, did reveal a greater variability of scores among the Weak Superego Group beyond the .05 level of confidence (using a conservative estimate) as compared to the Strong Superego Group.

The  $F$  test was also used to analyze the equality of variance of the total errors of the two groups on the different subtests of the Progressive Matrices (1938 and 1947) scales. The results revealed that the Weak Superego Group had a greater variability of scores than the Strong Superego Group as a result of having more errors on Set A (near the .01 level of confidence using a conservative estimate) and Set Ab (beyond the .05 level of confidence using a conservative estimate) of the Colored Matrices (1947) scale, and on Set D of the Matrices (1938) scale (beyond the .01 level of confidence using the conservative estimate).

### DISCUSSION

Significant differences between the Strong Superego and Weak Superego groups of delinquents were obtained, thus supporting the major hypothesis of this study, i.e., the Progressive Matrices test is sensitive in reflect-

ing differences in superego strength and functioning; a high score on the Matrices is associated with a measure of strong superego on the PAT, and a low score on the Matrices is associated with a measure of weak superego on the PAT when general ability is held constant.

There were consistent and reliable differences in performance between the two groups of delinquents on the Progressive Matrices scales supporting Hypothesis 1, i.e., delinquents who have more superego strength and control over impulse expression will obtain higher scores on the Progressive Matrices (Colored 1947 scale and 1938 scale, separately and combined), as compared to delinquents who have less superego strength and control over impulse expression.

In view of the results, it may be hypothesized that the "Strong Superego—Back to Work After Injury, Passivity, Fatigue, Sociophilia, and/or Anger" personality trait had a considerable influence on the test performance of one of the delinquent groups. Significant differences between the test scores of the Strong Superego and Weak Superego groups tend to support Johnson's (1953a, 1953b, 1953c) belief that an objective Raven score may provide some impression of superego capacity. The results of this study further support the possible use of the Progressive Matrices as a convenient screening test for assessing superego strength and functioning among delinquent boys of average intelligence who are committed to industrial schools. If institutionalized delinquents can be segregated on the basis of weak and strong superegos, perhaps different rehabilitation and education programs could be incorporated in helping them to adjust within the institution as well as in society (Aichorn, 1951).

It was observed that the Weak Superego Group of delinquents tended to be careless, haphazard, and impulsive in their performance on the Matrices as compared to the Strong Superego Group who seemed to be more meticulous in their performance. Porteus (1945) noticed that some of the delinquents were careful and exact in their performance on the Mazes, although most delinquents had decided tendencies toward

<sup>2</sup> "For data not satisfying the assumption of the  $F$  test, it would be far better to proceed with the  $F$  test and require that an obtained  $F$  reach the .01 level in order to be sure of significance at better than the .05 level or that an obtained  $F$  reach the .005 level for significance at near the .01 or .02 level [p. 362]."



recklessness and impulsiveness of behavior which nullified their planning actions.

The results of this study do show significant differences between the two groups in relation to length of time involved in test performance, thus supporting Hypothesis 2, i.e., delinquents who have more superego strength and control over impulse expression will be more ego involved with their performance on the combined Progressive Matrices (1938 and 1947) scales, and thus will take a longer time in their performance than delinquents who have less superego strength and control over impulse expression. The average performance time for the Strong Superego Group was a little over 26 minutes, and for the Weak Superego Group a little over 15 minutes. It would be interesting to note what changes would occur in the Matrices' scores, especially those of the Weak Superego Group, if both groups were asked to take a longer time in their performance and to recheck their work.

The Weak Superego Group's tendency to impulsive actions seemed to handicap their ability to reason and apprehend the governing principle in the Matrices' patterns. They appeared to be less ego involved with the test and tended to be more prone to guessing than the Strong Superego Group. The poor performance of the Weak Superego Group was noticed very early in the first two Sets A and Ab of the Colored Children's Matrices (1947) scale. Although they seemed to show some signs of ego involvement during the middle of their test performance, they soon lapsed into their old habits when the patterns became increasingly harder. This was especially noticed in Set D of the Matrices (1938) scale.

The test results also tend to corroborate Hypothesis 3, i.e., there will be a greater variability of scores on the combined Progressive Matrices scales among delinquents who have less superego strength and control over impulse expression than among delinquents who have more superego strength and control over impulse expression. Analysis of the test scores of the two groups revealed a greater variability resulting in more errors among the subjects in the Weak Superego Group than among the subjects in the Strong Superego

Group. Another possible explanation for the greater number of errors and guessing by the Weak Superego Group may be attributed to their chronically poor social adjustment. This may have handicapped their early schooling, and thus never allowed them the opportunity to set up standards of mature, independent self-accomplishment when faced with new learning situations.

The Henmon-Nelson Tests of Mental Ability have a reported validity coefficient around the .80s with other well-known intelligence tests (Buros, 1945). On the basis that the two groups of delinquents possess an average level of intelligence as measured by the Henmon-Nelson Tests of Mental Ability, a comparison of intellectual levels was made with the scores of the two groups using Raven's Progressive Matrices (1938) norms. According to Raven (1951), the Colored Children's Matrices (1947) scale (by omitting Set Ab) can be combined with the Matrices (1938) scale, Sets C, D, E, in order to use the norms for Progressive Matrices (1938).

The Matrices scores of the Strong Superego Group would classify two of the subjects as "definitely above average in intellectual capacity," six as "intellectually average," and four as "intellectually below average in intellectual capacity." In the Weak Superego Group, three of the subjects would be classified as "intellectually average," six as "definitely below average in intellectual capacity," and three as "intellectually defective."

The results of this study seem to support the findings of others that the Progressive Matrices test measures something other than general intelligence. Factor analytic studies of the Matrices, as well as batteries of reasoning tests, have found numerous factors besides "general reasoning" playing an important part in the actual solution of the test items. It seems to be the general consensus of many investigators that personality factors and emotional attitudes affect "clear thinking" in test performance. The test-retest reliability of the Progressive Matrices may not only be influenced by the nature of the test items themselves, but also by the emotional state of the subject taking the test at dif-

ferent intervals. It has been demonstrated that personality factors, such as superego strength, affect the Progressive Matrices as a valid measure of general intelligence and clear thinking.

Juvenile delinquents, as demonstrated in this study, seem to have more than one type of behavior and personality configuration. The two groups of delinquents in this study were selected to reveal different cleavages in ego and superego alignment described in the literature. Consistent and reliable differences of the two groups were also revealed in their behavior and adjustment records at the institution.

A comparison between the Strong Superego and Weak Superego groups of delinquents was made on the number of demerits they received during their first three months residence at the institution. An analysis of the total demerits received by the two groups (based on the average rating of the demerit system at the institution) by the Mann-Whitney Test revealed significant differences between the two groups beyond the .02 level of confidence. It was noted that during the first three months' residence at the institution, the Strong Superego Group had earned a mean of 4.83 demerits ( $SD = 5.26$ ), while the Weak Superego Group had earned a mean of 16.08 demerits ( $SD = 11.31$ ).

A further comparison of the two groups was made on the number of runaways from the institution during their first 3 months residence. Using a chi square test of "goodness of fit," it was noted that the Weak Superego was more likely to run away (beyond the .05 level of confidence) from the institution than the Strong Superego Group during its first 3 months period of residence. It is interesting to note that after the testing was completed, three of the boys in the Weak Superego Group did run away from the institution and were later apprehended. This may further relate to the Progressive Matrices and PAT as meaningful predictive tests of important behavior and adjustment patterns in the institution.

It was not within the scope of this study to cover the major sociocultural and environmental factors which have been reported to

contribute to delinquent behavior, because of the very small sampling of the delinquent population. However, regardless of sociocultural and environmental factors, the Progressive Matrices scales are sensitive in reflecting consistent and reliable differences in superego strength and functioning in a delinquent population of boys when general ability is held constant.

Although there is a growing awareness of the need to establish diagnostic facilities and predictive tools for the youthful offender, generalizations from the results of this study should be limited only to similar groups of institutionalized delinquents with caution. Further studies are in order with respect to the use of the Progressive Matrices as a predictive tool in observing and classifying delinquent behavior, as well as behavior on a large scale.

#### REFERENCES

- AICHORN, A. *Wayward youth*. (10th ed.) New York: Viking Press, 1951.
- BOLIN, B. A comparison of Raven's Progressive Matrices (1938) with the ACE Psychological Examination and the Otis Gamma Mental Ability Test. *J. consult. Psychol.*, 1955, 19, 400.
- BROMLEY, D. B. Primitive forms of response to the Matrices Test. *J. ment. Sci.*, 1953, 99, 374-393.
- Buros, O. K. (Ed.) *The nineteen forty mental measurements yearbook*. Arlington: Gryphon Press, 1945.
- DESAI, M. The relationship of the Wechsler-Bellevue Verbal Scale and the Progressive Matrices Test. *J. consult. Psychol.*, 1955, 19, 60.
- ESSLER, K. R. (Ed.) *Searchlights on delinquency*. New York: International Universities Press, 1949.
- FRIEDLANDER, KATE. *The psychoanalytic approach to juvenile delinquency*. New York: International Universities Press, 1947.
- FROSCHÉ, J. (Ed.) *The annual survey of psychoanalysis*. 1952, 1953. New York: International Universities Press, 1956. Pp. 104-185.
- HALL, JULIA C. Correlation of a modified form of Raven's Progressive Matrices (1938) with the Wechsler Adult Intelligence Scale. *J. consult. Psychol.*, 1957, 21, 23-26.
- HALSTEAD, H. An analysis of the Matrix (Progressive Matrices) test result on 700 neurotic (military) subjects, and a comparison with the Shipley Vocabulary Test. *J. ment. Sci.*, 1943, 89, 202-215.
- HORA, T. The dissocial superego. *Amer. J. Psychother.*, 1952, 6, 513-519.
- JOHNSON, ELIZABETH. The clinical use of Raven's Progressive Matrices to appraise potential for progress in play therapy: A study of institution-

- alized mentally and educationally retarded children. *Amer. J. Orthopsychiat.*, 1953, 23, 391-398.
- JOHNSON, ELIZABETH. Individual patterns of emotional functioning in children of comparable I.Q.'s—Implications for education. *Amer. J. ment. Defic.*, 1953, 57, 681-686.
- JOHNSON, ELIZABETH. Klopfer's prognostic scale used with Raven's Progressive Matrices in play therapy prognosis. *J. proj. Tech.*, 1953, 17, 320-326.
- LEVINE, B., & ISCOE, IRA. A comparison of Raven's Progressive Matrices (1938) with a short form of the Wechsler-Bellevue. *J. consult. Psychol.*, 1954, 18, 10.
- MCNEMAR, Q. On Wilson's distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1957, 54, 361-362.
- MARTIN, A. W., & WIECHERS, J. E. Raven's Colored Progressive Matrices and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1954, 18, 143-144.
- OBERNDORF, C. P. Psychotherapy of work. *Bull. Menninger Clin.*, 1951, 15, 77-84.
- PORTEUS, S. D. Q-scores, temperament, and delinquency. *J. soc. Psychol.*, 1945, 21, 81-103.
- RAVEN, J. C. *Guide to using Progressive Matrices* (1947) *Sets A, Ab, B.* (2nd ed.) London: H. K. Lewis, 1951.
- RAVEN, J. C. *Guide to using Progressive Matrices* (1938). (5th ed.) London: H. K. Lewis, 1954.
- REDL, F., & WINEMAN, D. *The aggressive child.* Glencoe, Ill.: Free Press, 1957.
- SLATER, P. Comment on "the conceptual assessment of intellectual ability." *Brit. J. Psychol.*, 1948, 39, 20-21.
- STACEY, C. L., & CARLETON, F. O. The relationship between Raven's Colored Progressive Matrices and two tests of general intelligence. *J. clin. Psychol.*, 1955, 11, 84-85.
- STACEY, C. L., & GILL, MARIE R. The relationship between Raven's Colored Progressive Matrices and two tests of general intelligence for 172 subnormal adult subjects. *J. clin. Psychol.*, 1955, 11, 86-87.
- TOMKINS, S. S., & MINER, J. B. *The Tomkins-Horn Picture Arrangement Test.* New York: Springer, 1957.
- WATSON, D. The validity and interchangeability of Terman-Merrill and Matrices data. *Brit. J. educ. Psychol.*, 1955, 25, 190-194.

(Received January 25, 1963)



## FACTORS OF PSYCHOPATHOLOGY IN THE WARD BEHAVIOR OF ACUTE SCHIZOPHRENICS<sup>1</sup>

ALLEN RASKIN

*Georgetown University*

AND

DEAN J. CLYDE

*George Washington University*

2 factor analyses were performed on nurses' ratings of the ward behavior of 417 acute schizophrenic patients. The 10 psychotic syndrome scores which Lorr, McNair, Klett, and Lasky (1962) derived from behavior observable in an interview were included in one of the analyses. 11 factors labeled Self-Care, Social Participation, Extraversion, Irritability, Anxiety, Guilt Feelings, Depression, Feelings of Unreality, Slowed Speech and Movements, Paranoid Projections, and Excitement were extracted from the ward behavior ratings. There was good correspondence between the factors of psychopathology extracted from patient ward behavior and the Lorr et al. psychotic syndromes.

Lorr, McNair, Klett, and Lasky (1962) present evidence for 10 postulated psychiatric symptom syndromes based on behavior observable in an interview. As noted by these authors, quantified indices of symptom syndromes have practical value as criteria of change in studies on effects of various treatment procedures and may also provide a useful approach to the hoary problem of patient classification. However, the symptom syndromes identified by Lorr et al. have certain limitations both as measures of improvement and as a schema for classifying psychiatric patients. First, many of the psychiatric rating scale items, which comprise the raw data for the Lorr et al. study, describe behaviors indicative of severe psychopathology; mild and moderate aspects of psychopathology are less well represented in these scales. Secondly, the behaviors sampled are limited to those that can be observed in an interview of approximately an hour's duration.

In the present study an effort was made to identify factors of psychopathology in the ward behavior of acute schizophrenics. Behaviors characterizing mild, moderate, and severe psychopathology are equally represented in the ward behaviors sampled. The 10 psychotic syndrome scores, derived from Form Number 3 of the Inpatient Multidimensional Psychiatric Scale (IMPS), were in-

cluded as marker variables in one of the factor analyses of patient ward behavior. The syndrome scores were included to clarify the meaning of emerging factors of psychopathology in patient ward behavior and to extend the meaning of the Lorr et al. symptom syndromes by identifying their correlates in patient ward behavior.

### METHOD

#### *Subjects*

The present study is part of a larger collaborative study of the effects of phenothiazine treatment on acute schizophrenic psychoses. The study sample consisted of 185 male and 232 female patients drawn from the psychiatric populations of two large, metropolitan receiving hospitals,<sup>2</sup> four state hospitals,<sup>3</sup> and three private institutions.<sup>4</sup> Newly admitted patients, 16 to 45 years of age, with no previous psychiatric hospitalization within the past 6 months and no evidence of childhood schizophrenia, mental deficiency, chronic or acute brain syndrome, epilepsy, drug addiction, or chronic alcoholism were included in the study. To satisfy the diagnosis of acute schizophrenic psychosis, patients had to manifest symptoms in at least two of the following categories of psychopathology: (a) thinking and speech disturbance, (b) catatonic motor behaviors, (c) paranoid ideation and be-

<sup>2</sup> D. C. General Hospital, Washington, D. C., and Malcolm Bliss Mental Health Center, St. Louis, Missouri.

<sup>3</sup> Boston State Hospital, Boston, Massachusetts, Kentucky State Hospital, Danville, Kentucky, Rochester State Hospital, Rochester, New York, and Springfield State Hospital, Sykesville, Maryland.

<sup>4</sup> Institute of Living, Hartford, Connecticut, Mercy-Douglass Hospital, Philadelphia, Pennsylvania, and Payne-Whitney Clinic, New York City.

<sup>1</sup> This research was supported by funds made available by Research Grant MYP4663 of the Psychopharmacology Service Center, National Institutes of Health, Public Health Service, United States Department of Health, Education and Welfare.

havior, (d) hallucinations, (e) delusional thinking, (f) disturbances of affect and emotion, and (g) disturbances of social behavior and interpersonal relations.

### Procedure

Before their second week in the hospital and prior to their assignment to one of the four drug treatment groups, patients were interviewed by two psychiatrists or a psychiatrist and a psychologist. Immediately following this interview, each interviewer rated the patient on the IMPS. In addition, two nurses rated each patient on the Ward Behavior Rating Scale (WBRS) (Burdock, Hardesty, Hakerem, & Zubin, 1960) at the end of the patient's first week of hospitalization and prior to his assignment to one of the drug treatment groups.

Form Number 3 of the IMPS consists of 46 9-point and 13 5-point unipolar rating scales. In addition, 19 checklist items are rated present or absent. Each of the 10 syndrome scores is represented in the rating schedule by 5 to 11 scale variables. The 10 syndrome scores are labeled Excitement, Paranoid Projections, Motor Disturbances, Hostile Belligerence, Depressive Agitation, Perceptual Distortion, Apathy and Retardation, Grandiose Expansiveness, Thinking Disorganization, and Disorientation.

The WBRS contained the original 150 check list items by Burdock et al. (1960) plus three additional items contributed by the staff of the Psychopharmacology Service Center, National Institute of Mental Health. Each check list item is rated present or absent.

As available electronic computer programs for a factor analysis were limited to 100 variables, it was not possible to include the 153 WBRS items and the 10 IMPS factor scores in a single analysis. Selection of WBRS items for inclusion in the analysis on the basis of rationally derived syndromes appeared to offer the best approach for identifying the major factors of psychopathology in the WBRS. However, as it is anticipated WBRS factor scores will be used as criterion measures in studies of drug and other treatment effects, there was also merit in using item reliability as a criterion for item selection. Consequently two factor analyses were performed.

Prior to the first analysis, the 153 WBRS items were rationally grouped by the authors and 14 syndromes of psychopathology identified. Five clinical psychologists, working independently, then assigned each of the WBRS items to one of the 14 rationally derived syndromes. The 90 WBRS items included in the first factor analysis met the criterion of agreement between four of the five judges on placement in a syndrome and in addition described behaviors rated as present in at least 10% of the study sample. The 10 IMPS syndrome scores for each patient were also included in this first analysis. The second factor analysis was based on

100 WBRS items meeting the criterion of an intraclass reliability coefficient of .35 or higher between the paired nurses ratings. Twenty-seven items failed to meet the criteria for inclusion in either analysis.

The dichotomous and continuous variables in the first analysis were intercorrelated by the Pearson product-moment, point biserial, and phi coefficient methods.<sup>6</sup> The correlation matrix for the dichotomous variables in the second analysis consisted of phi coefficients. Both correlation matrices were factor analyzed using Hotelling's principal components method with unity in the diagonals.<sup>6</sup> A normal varimax rotation, which provides an orthogonal solution, was then performed on all factors with eigenvalues of one or greater.

### RESULTS

Eleven factors containing at least three significant WBRS items were extracted from both factor analyses. Item significance was arbitrarily defined as a loading of .40 or higher on one factor and no loading of .40 or higher on any other factor. As the IMPS syndrome scores are based on a weighted combination of at least five individual scales, were derived from an oblique rotation and provide useful frames of reference for the WBRS factors, they are included in a WBRS factor if they load .27 or higher on that factor. With minor exceptions the same WBRS factors emerged from both factor analyses despite the previously described dif-

<sup>6</sup> Although the point biserial and phi coefficient are systematically lower than the respective Pearson  $r$  equivalents, the biserial correlation coefficient and the tetrachoric correlation coefficient, they do not make the assumptions that the dichotomous variable(s) is actually continuous and its underlying distribution is normal bivariate. As many of the WBRS items fail to meet these assumptions, the point biserial and phi coefficient methods were chosen.

<sup>6</sup> Unity was chosen as the diagonal entry for two reasons. First, in an empirical study, Tyler and Michael (1958) compared the factor structure of anthropometric data when unities and communality estimates were used. The same common factors were identified for both types of diagonal entry with the degree of similarity of loadings on the same factors exceedingly high as shown by two different indices of similarity. Secondly, use of unity as the diagonal entry permits derivation of exact factor scores for individual patients whereas with communality estimates approximate factor scores are obtained.

TABLE 1  
FACTOR A: SELF-CARE

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
9	Needs help in dressing	-.75	-.72
10	Follows instructions	.60	.56
28	Repeats words and phrases	-.44	
34	Answers questions about himself		.41
36	Takes care of personal appearance	.76	.78
42	Is able to make own bed	.69	.68
46	Knows new and old personnel	.40	
53	Does chores about ward	.42	.42
54	Makes a nuisance of himself	-.42	-.40
60	Gets along with other patients	.52	.49
65	Has to be helped in the bathroom	-.61	-.59
72	Speech is sensible and connected	.58	.50
93	Refuses things expected of him	-.68	-.68
95	Looks sloppy	-.66	-.69
98	Has to be helped to stick to things	-.42	-.41
100	Keeps self clean	.77	.79
102	Washes or showers himself	.78	.80
109	Cooperates with hospital personnel	.61	.64
120	Behaves acceptably off ward		.59
IMPS	Disorientation	-.62	
IMPS	Motor Disturbances	-.68	
IMPS	Thinking Disorganization	-.42	

<sup>a</sup> Percentage of variance, 10.5; first factor analysis.

<sup>b</sup> Percentage of variance, 9.5; second factor analysis.

ference in criteria for item inclusion.<sup>7</sup> Consequently, combined results for both analyses are reported.

Factor A (see Table 1) contains the largest number of significant items and also accounts for the greatest percentage of total item variance. This factor, labeled Self-Care, reflects the patient's ability to care for his hygienic needs, personal appearance, and daily ward chores. Conversely, this factor is indicative of the extent of nursing care and attention required by the patient. Three IMPS syndromes, Disorientation (e.g., disoriented as to hospital, age, year), Motor Disturbances (e.g., talks to self, slovenly appearance, repetitive movements), and Thinking Disorganization (e.g., neologisms, stereo-

typed speech, incoherent answers) have significant loadings on Factor A. For the most part, these IMPS syndromes have in common behaviors characterizing severe loss of contact with reality. Hence, as one might expect, the patient's inability to care for his personal needs and the extent of his demands on the time and services of nursing personnel is inversely related to degree of contact with reality. From a nursing standpoint, other aspects of psychopathology such as hostility, paranoid ideation, and depression are less closely associated with patient self-care activities.

Factor B, Social Participation, is a measure of the patient's interest in being with other patients and extent of participation in social and recreational activities. In terms of psychopathology, it is the converse of social isolation and withdrawal. As the IMPS scales do not sample patient behaviors in this area, there is no IMPS syndrome comparable to this factor.

<sup>7</sup> Identification of the same common WBRS factors in both analyses and the high degree of similarity of loadings on the same factors for both analyses indicate little resultant distortion of the first factor analysis as a consequence of analyzing point biserial correlations and phi coefficients in the presence of product-moment correlations.



TABLE 2

## FACTOR B: SOCIAL PARTICIPATION

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
25	Is friendly with someone on ward	-50	-47
26	Jokes with fellow patients	-73	-74
39	Makes small talk		-62
45	Joins in social games	-71	-74
57	Shows interest in going to O.T.		-61
58	Likes to talk about his interests	-62	
66	Shows pleasure in recreation	-70	-80
69	Shows interest in outside events	-56	-57
117	Takes part in ward activities		-76
121	Talks over happenings on ward		-63
124	Likes to be occupied		-69

<sup>a</sup> Percentage of variance, 4.9; first factor analysis.

<sup>b</sup> Percentage of variance, 8.1; second factor analysis.

Although Factor C, Extraversion, is rationally related to Social Participation, it describes a more active and assertive social role with the patient initiating conversation and volunteering information about himself.

The IMPS Hostile Belligerence syndrome is represented in the WBRs by Factor D, Irritability. This factor characterizes an attitude of discontent and sullenness. Motor expressions of hostility such as "Tears or breaks things" and "Hits others" do not load on this factor. The Hostile Belligerence syndrome on the IMPS is similarly comprised of items reflecting hostility in attitude and verbal expression.

In contrast to the IMPS, a separate Anxiety factor, Factor E, emerges from the WBRs items. In addition to observable signs of anxiety or distress this factor also describes certain behavioral accompaniments of a heightened anxiety state such as hesi-

TABLE 3

## FACTOR C: EXTRAVERSION

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
8	Volunteers information		-42
51	Is silent when relatives visit	59	
63	Strikes up conversation		-45
73	Likes to be with the opposite sex	-57	-59

<sup>a</sup> Percentage of variance, 1.8; first factor analysis.

<sup>b</sup> Percentage of variance, 2.3; second factor analysis.

TABLE 4

## FACTOR D: IRRITABILITY

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
7	Gets angry when kidded		72
14	Has temper tantrums		44 56
16	Is irritable and grouchy		70 70
23	Is impatient		66
27	Gets angry when questioned		70 70
49	Shows frequent changes of mood		42 49
74	Is easily upset		63
119	Accepts hospital routine	-54	
136	Makes sarcastic remarks	63	62
146	Has an angry expression	70	70
IMPS Hostile Belligerence			41

<sup>a</sup> Percentage of variance, 5.0; first factor analysis.

<sup>b</sup> Percentage of variance, 5.4; second factor analysis.

tancy in making up one's mind and an appearance of helplessness and perplexity. The IMPS has three anxiety items. Two relate to verbal reports of anxiety in anticipation of future problems and in anticipation of vague indefinable future misfortunes. In the study by Lorr et al. (1962) these items had significant loadings on the IMPS Agitated Depression syndrome. The third IMPS anxiety item, "Manifests overt signs of tension," did not have a significant loading on this syndrome. Perhaps the patient who is both anxious and depressed verbalizes feelings of anxiety to the interviewer whereas many anxious patients lacking depressive features do not.

The IMPS Depressive Agitation syndrome had high loadings on two factors, Guilt Feelings (Factor F) and Depression (Factor G). Many of the behaviors in the IMPS Depressive Agitation syndrome are represented in

TABLE 5

## FACTOR E: ANXIETY

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
18	Is tense and anxious		-61
29	Appears helpless or perplexed	-40	-63
48	Is fidgety and nervous	-70	
114	Is hesitant in making up mind		-71

<sup>a</sup> Percentage of variance, 2.4; first factor analysis.

<sup>b</sup> Percentage of variance, 2.6; second factor analysis.

TABLE 6  
FACTOR F: GUILT FEELINGS

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
35	Loses things		-45
77	Talks about his unworthiness	-77	-49
78	Mixes up words		-45
140	Says he is no good	-77	
IMPS	Depressive Agitation	-41	

<sup>a</sup> Percentage of variance, 1.9; first factor analysis.

<sup>b</sup> Percentage of variance, 1.6; second factor analysis.

Factors F and G of the WBRs, i.e., depressed mood, guilt and remorse, ideas of sinfulness, and recurring thoughts.

The Grandiose Expansiveness syndrome had a significant loading on WBRs Factor H, Feelings of Unreality. Both factors describe delusions of unusual power or an outside force controlling or directing one's behavior.

The IMPS Apathy and Retardation syndrome is represented on the WBRs by Factor I, Slowed Speech and Movements. These factors have in common retardation in motor behavior and in speech.

IMPS Paranoid Projections had its highest loading (.27) on WBRs Paranoid Projections (Factor J). However, in contrast to the WBRs items, the IMPS scales tend to describe a higher level of psychopathology, e.g., ideas of reference, ideas of persecution, and ideas of conspiracy. This difference in level of psychopathology may account for the relatively low loading of IMPS Paranoid Projections on the comparable WBRs factor.

TABLE 7  
FACTOR G: DEPRESSION

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
6	Looks sad	76	77
138	Speaks in a sad voice	72	80
147	Says he feels blue	67	
IMPS	Depressive Agitation	40	

<sup>a</sup> Percentage of variance, 2.7; first factor analysis.

<sup>b</sup> Percentage of variance, 2.8; second factor analysis.

TABLE 8  
FACTOR H: FEELINGS OF UNREALITY

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
126	Claims he is being controlled	72	77
135	Says things have special meanings	48	71
142	Says that all or part of himself is outside his body, or that someone is in his body		57 49
IMPS	Grandiose Expansiveness		56

<sup>a</sup> Percentage of variance, 2.8; first factor analysis.

<sup>b</sup> Percentage of variance, 2.2; second factor analysis.

The final WBRs factor, Excitement (Factor K), is rationally related to IMPS Excitement. Both contain items sampling overtalkativeness and braggartism. However, the three WBRs items in the Excitement factor were not included in the first factor analysis containing the IMPS syndrome scores. Consequently, the loading of IMPS Excitement on the comparable WBRs factor is not known.

The only IMPS syndrome having no apparent counterpart in the WBRs factors is Perceptual Distortion. This syndrome is comprised of scales which describe visual and auditory hallucinations. Comparable items are included in the WBRs but most were excluded from the factor analysis containing the psychotic syndrome scores because they occurred with less than 10% frequency.

In sum, despite differences in behaviors sampled (The IMPS samples patient interview behavior whereas the WBRs is based on patient ward behavior) and level of psychiatric sophistication of individuals rating

TABLE 9  
FACTOR I: SLOWED SPEECH AND MOVEMENTS

Number	Item content	Factor loading	
		I <sup>a</sup>	
67	Is slow in his movements	63	
112	Voice is flat and monotonous	50	
116	Speaks in slow, drawn out manner	74	
IMPS	Apathy and Retardation	38	

<sup>a</sup> Percentage of variance, 2.7; first factor analysis.

TABLE 10  
FACTOR J: PARANOID PROJECTIONS

Number	Item content	Factor loading	
		I <sup>a</sup>	II <sup>b</sup>
85	Says people hate him	70	
89	Says others want to hurt him	66	
125	Says people dislike him	83	76
IMPS	Paranoid Projections	27	

<sup>a</sup> Percentage of variance, 2.5; first factor analysis.  
<sup>b</sup> Percentage of variance, 1.4; second factor analysis.

TABLE 11  
FACTOR K: EXCITEMENT

Number	Item content	Factor loading
		II <sup>a</sup>
82	Brags or boasts	46
108	Talks and talks	62
111	Is constantly moving about	54

<sup>a</sup> Percentage of variance, 2.4; second factor analysis.

the IMPS and WBRs, there is good correspondence between the IMPS psychotic syndromes and the WBRs factors. In varying degrees, 8 of the 10 psychotic syndromes have counterparts in the WBRs factors. The identification of correlates of the psychotic syndromes in patient ward behavior clarifies the meaning of these syndromes and also lends support to their use as a schema for classifying psychiatric patients. Further, the practical value of psychotic syndromes as change measures may be enhanced by combining the psychotic syndrome scores with comparable factor scores from the WBRs. Finally, the study has demonstrated that there are meaningful factors of psychopathology in patient ward behavior. The latter finding suggests that current use of the WBRs to obtain a total morbidity score is an uneconomical use of this instrument.

Utilization of separate WBRs factors as change measures may provide more effective means of discriminating the effects of various treatment procedures. The separate WBRs factor scores may also be useful for describing or categorizing nuances of patient ward behavior not possible with a single morbidity measure.

#### REFERENCES

- BURDOCK, E. I., HARDESTY, A. S., HAKEREM, G., & ZUBEN, J. A ward behavior rating scale for mental patients. *J. clin. Psychol.*, 1960, 16, 246-247.  
 LORR, M., MCNAIR, D. M., KLETT, C. J., & LASKY, J. J. Evidence of ten psychotic syndromes. *J. consult. Psychol.*, 1962, 26, 185-189.  
 TYLER, F., & MICHAEL, W. B. An empirical study of the comparability of factor structure when unities and communality estimates are used. *Educ. psychol. Measmt.*, 1958, 18, 347-354.

(Received January 28, 1963)



## INTERPERSONAL ORIENTATION IN RELATION TO HYPNOTIC SUSCEPTIBILITY<sup>1</sup>

P. M. BENTLER

*Stanford University*

2 samples of women ( $N = 37$ ,  $N = 47$ ) described themselves on the Interpersonal Check List (ICL) and were hypnotized according to the Stanford scales, Forms A and C. One sample of 43 men described themselves on the ICL; these men were hypnotized according to procedures of Form A. The Docile-Dependent, Cooperative-Overconventional, and Responsible-Hypernormal octants showed the highest correlations with hypnotizability. ICL scores based on factor analyses showed that a "positive" and affiliative interpersonal orientation correlated positively with hypnotizability, while the "negative" orientation factor and the dominance-submission factor were unrelated to hypnotizability. A comparison of correlations of hypnotic measures with ICL scores indicated that the correlations which approached or reached significance for the combined female samples were almost uniformly smaller for Form C than for Form A.

Most researchers in the field of hypnosis feel that the personalities of highly susceptible hypnotic subjects differ from the personalities of refractory subjects, yet the nature of this difference eludes precise quantitative measurement. Hilgard and Lauer (1962), for example, found none of the 18 California Psychological Inventory (CPI) (Gough, 1957) scales to yield correlations with hypnotizability which replicated significantly across sexes. In samples with more than 100 subjects per sex, only three out of 36 correlations were higher than .20, and the highest was .26. An item analysis was not much more successful. Only 33 items for males and 36 items for females out of the 480 items of the CPI discriminated the upper half from the lower half of hypnotic subjects. Scales constructed of these items correlated in the .60s on the sample from which the data came, but on cross validation these correlations fell to less than .10. Weitzenhoffer

and Weitzenhoffer (1958) administered the Guilford-Zimmerman Temperament Survey (1949) and the 16 P.F. questionnaire (Cattell, Saunders, & Stice, 1953) to 200 subjects. The highest of 24 obtained correlations was .18. When significant results are reported, the prognosis is generally not much more favorable. Replicating significant results is quite difficult. Furneaux and Gibson (1961) and Lang and Lazovik (1962) have, for example, reported data indicating a negative correlation between the Maudsley Personality Inventory (Bendig, 1959; Eysenck, 1959) measure of neuroticism and susceptibility to hypnosis. Hilgard and Bentler (1963), on the other hand, found a correlation of .00 with 143 subjects.

The study reported here concerns itself with the personality characteristics of "basic trust," as described by Erikson (1950). Trust is essentially an attitude toward oneself and the world, and it is basic in the sense that it appears early in life and need not be conscious. This variable is theoretically related to hypnotizability since hypnosis involves relinquishing control to another person. In his developmental-interactive theory of hypnosis, Hilgard (1962) points out that the love relationship between parent and child is an antecedent which may be developmentally related to hypnotizability through basic trust. Furthermore, this type of trust

<sup>1</sup>From the Laboratory of Human Development. This investigation was supported in part by a Public Health Service fellowship number MPM-15, 840 from the National Institute of Mental Health, United States Public Health Service. The hypnotic scores were obtained within the investigations supported by a grant from the National Institute of Mental Health (Grant M-3859) to Ernest R. Hilgard. I am indebted to Nancy Wiggins, Kenneth M. Colby, Ernest R. Hilgard, Perry London, and André M. Weitzenhoffer for their constructive criticisms of the manuscript.

has implications for the transference aspects of hypnosis.

Basic trust, of course, cannot be measured easily. Only superficial manifestations of this trait could be assessed. Barber (1960) attempted to construct a questionnaire measure of this trait from the item pool of the Guilford-Zimmerman Temperament Survey (1949) and the Webster-Sanford-Freedman (1955) *F* Scale. This scale differentiated the one refractory subject out of 20 hypnotizable subjects, but apparently it has not held up under cross-validation (Barber & Glass, 1962).

Basic trust may be related to hypnosis by assuming that trust established early in life has residues which are reflected in a person's interpersonal orientation. Recent factor analyses performed at Stanford (Wiggins, 1961) of an existing instrument indicates that the Leary Interpersonal Check List (LaForge & Suczek, 1955) contains a "Love" factor which appears to reflect affiliation, approval seeking, friendliness, and sociableness which may be considered a "positive" interpersonal orientation consonant with basic trust.<sup>2</sup> Since the "Love" factor is theoretically linked to hypnotizability, the existing instrument was chosen as a first approximation in evaluating the relevance of this type of orientation for hypnosis. If the results appear promising a next step would be to construct an instrument measuring this factor more adequately.

## METHOD

### *Interpersonal Check List*

The Interpersonal Check List (ICL) is a self-rating adjective-type check list developed by LaForge and Suczek for inclusion in the Leary (1957) *Interpersonal Diagnosis of Personality*. The ICL can measure more than one "level" of personality, but Leary's Level II, the level of conscious description, is generally used in conjunction with the ICL. This study is concerned with that level. There are 8 octants, or variables, in the ICL, each of which is represented by 16 items. The items were selected on the basis of their rational content, intensity ratings, and endorsement frequency. The octants are generally identified by two alphabetical letters and

are descriptively summarized as follows: AP = managerial-autocratic; BC = competitive-narcissistic; DE = aggressive-sadistic; FG = rebellious-distrustful; HI = self-effacing-masochistic; JK = docile-dependent; LM = cooperative-overconventional; NO = responsible-hypernormal. A person checking some of the items in each octant is considered normal, while a person checking most or all of the items in an octant has unhealthy amounts of the trait. For example, being managerial in orientation is fine, but having too much of this trait shows undesirable tendencies to being autocratic.

Although the ICL is usually described in terms of two bipolar dimensions named Love-Hate and Dominance-Submission, recent factor analyses by Wiggins (1961) indicate that the raw scores are better described by three factors. Wiggins (1961) performed a Thurstone centroid method of factor extraction and rotated to visual criteria of simple structure. Although the factor structure differed slightly for the sexes, three orthogonal factors emerged: Love, Hate, and a bipolar Dominance-Submission factor. A principal axes method with Varimax rotation was used by Wiggins and Bentler on the same data, and three orthogonal factors were found again.<sup>2</sup> These three factors will be described briefly since they are used in this study. Factor I is Leary's dominance-submission bipolar factor. Using a loading of .32 to define a significant factor loading, octants AP, BC, and DE loaded positively and HI loaded negatively on this factor for both sexes. The highest positive loading was on the Managerial-Autocratic octant for both sexes. Factor II represents a positive interpersonal orientation and has been called Love. It loads significantly on octants HI, JK, LM, and NO for both sexes. The highest loading is on the Cooperative-Overconventional octant for both sexes. Factor III represents a negative interpersonal orientation for both sexes; it loads highest on the "hate" side of the Leary circle, though the significant loadings are not identical for the sexes. For females, Factor III loads on DE, FG, HI, and JK and for males on BC, DE, FG, and HI. For both sexes, however, the highest loading is on the Rebellious-Distrustful octant.

### *Factor Measurement and Hypotheses*

As was pointed out in the introduction, a positive interpersonal orientation is hypothesized to be related to hypnotic susceptibility. The ICL octants comprising the trusting, cooperative orientation should correlate highest with hypnotizability (HI, JK, LM, NO) and Factor II scores should correlate positively with hypnotizability. Factor scores were arrived at for each person by summing scores on octants on which the loadings for a factor were over .32. Octants were given equal positive or negative weight, depending on the sign of the loading, but irrespective of the size of the loading as long as it was over .32. Thus, subjects' scores on Factor II were computed by summing their item endorsements of octants HI, JK, LM, and NO.

<sup>2</sup> Wiggins, Nancy, & Bentler, P. M. The structure of the Interpersonal Check List. Unpublished manuscript, 1962.



TABLE 1  
CORRELATIONS BETWEEN HYPNOTIZABILITY AND ICL SELF AND FACTOR SCORES

Sample (sex) N	Form A				Form C		
	I(F) 37	II(F) 47	I+II(F) 84	III(M) 43	I(F) 37	II(F) 47	I+II(F) 84
<i>ICL Octant</i>							
AP Managerial-Autocratic	-.04	.23	.11	-.04	-.07	.23	.12
BC Competitive-Narcissistic	-.24	.11	-.03	.12	-.20	.21	.05
DE Aggressive-Sadistic	-.12	-.14	-.13	-.02	-.13	-.08	-.10
FG Rebellious-Distrustful	-.01	-.19	-.11	-.08	.14	-.22	-.06
HI Self-effacing-Masochistic	.08	.13	.10	.04	.15	.10	.11
JK Docile-Dependent	.18	.25	.22*	.23	.25	.14	.18
LM Cooperative-Overconventional	.40**	.30*	.34***	.24	.39**	.22	.29***
NO Responsible-Hypernormal	.26	.25	.24*	.20	.22	.22	.21
<i>ICL Factors</i>							
I. AP+BC+DE+HI	-.19	.03	-.06	-.10	-.20	.10	-.01
II. HI+JK+LM+NO	.31	.32*	.31***	.24	.35*	.25	.28**
III. Women: DE+FG+HI+JK	.05	.08	.06		.16	.01	.07
III. Men: BC+DE+FG+HI				.03			

\*  $p < .05$ .\*\*  $p < .02$ .\*\*\*  $p < .01$ .

Factor III, which represents a negative interpersonal orientation, might be negatively related to hypnotizability. The method of computing factor scores, however, precludes this being a reasonable hypothesis. An inspection of the factor loadings reveals that for women Factor III has loadings on the submission and dependent quadrants HI and JK; since these scores enter as equally into the factor score as aggressive or rebellious type items, the negative correlation which could otherwise be expected would be minimized. This overlap in fact produces a good-sized correlation between Factor II and III scores in this sample. In two samples of women to be described this correlation was .51 and .58, respectively. For a sample of men, the Factor II and III scores correlated .22. This correlation is lower since only one octant is common to both factors. Since both a positive (II) and a negative (III) correlation with a criterion is not to be expected between variables which are themselves positively correlated, the net expected effect would be to attenuate the Factor III correlation with hypnotizability.

Factor I, the dominance-submission factor, could be linked to hypnotizability by considering susceptible subjects as tending to "submit" to the hypnotist while unsuceptible subjects tend to do battle with the authority represented by the hypnotist. White (1937) has advanced the hypothesis that active hypnotic subjects are judged higher in the trait of deference, and he has presented evidence to substantiate this claim. These are the type of subjects generally seen at universities (Lang & Lazovik, 1962). From this view, Factor I should correlate negatively with hypnotizability.

### Subjects and Procedure<sup>3</sup>

For purposes of deriving the factors described above, Wiggins (1961) administered the ICL to 235 men and 154 women in introductory psychology classes under standard instructions to rate themselves on the 128 items. Three separate samples of subjects were drawn from the same introductory classes and were hypnotized according to the procedures of the Stanford Hypnotic Susceptibility Scale, Form A (Weitzenhoffer & Hilgard, 1959). Two of the samples consisted of women ( $N=37$ ;  $N=47$ ), and one of men ( $N=43$ ). The women were hypnotized a second time according to Form C of the Stanford Scale (Weitzenhoffer & Hilgard, 1962). This scale contains more cognitive items such as a dream, an age regression, and hallucinations. These subjects were all "coerced volunteers" (Boucher & Hilgard, 1962), and, as such, their hypnosis scores were in accord with the reported norms for the Stanford Scales (Hilgard, Weitzenhoffer, Landes, & Moore, 1961; Weitzenhoffer & Hilgard, 1962).

<sup>3</sup> The author is grateful for the generous cooperation of various persons. While the entire staff and assistants of the Laboratory of Human Development carried out the hypnotic sessions as part of the regular research program, specific mention should be made of Nancy Wiggins and Alvin I. Haimson who made the ICL scores available. Special thanks go to Forrest Young for helping with the statistical analyses.



## RESULTS

The correlations between hypnotic susceptibility and the various ICL measures are reported in Table 1. As can be seen the octants reflecting affiliation, approval seeking, and a generally conventionally accepting attitude correlate highest with hypnosis. For the two samples the Cooperative-Overconventional octant LM exhibits the highest correlations with hypnotizability in these samples. These correlations are the only ones that replicate significantly for two samples separately ( $r = .40$ ;  $r = .30$ ). None of the correlations in the male sample is statistically significant, though their pattern is identical to the two female samples' correlations. The octant correlations are about the same for Form A as for Form C in the female samples. However, the correlations which approach or reach significance for the combined samples are almost uniformly smaller for Form C than Form A.

Of the three factors, only Factor II correlations approach significance. It is not significant for males ( $r = .24$ ), is significant for Form C in the first female sample ( $r = .35$ ), and is significant for Form A in the second female sample ( $r = .32$ ). For the two female samples combined, the correlation is highly significant for both Forms A and C.

## DISCUSSION

Although the obtained correlations are not large, they do replicate over several samples. The consistency of the correlation between hypnotizability and a positive interpersonal orientation (Factor II) suggests that here is something upon which to build. No contradictory results have yet appeared in the literature. More importantly, the results make theoretical sense in light of the developmental-interactive propositions outlined by Hilgard (1962).<sup>4</sup>

<sup>4</sup> During the time this paper has been in press, Barber and Calverley attempted to replicate the present findings. They were unsuccessful (personal communication). There were a variety of differences between the replication and the procedures of the present study. The hypnotic scales, the ICL format, the subjects' instructions and prehypnotic experiences, and the experimenters were some of these.

In a recent study (Lang & Lazovik, 1962) the Edwards Personal Preference Schedule (EPPS) (1953) was administered to 32 subjects. The scale correlating most highly with hypnotizability was the Affiliation Scale ( $r = .36$ ). This is clearly in line with the present results. The only other EPPS scale showing a positive (nonsignificant) correlation with hypnotizability was the Deference Scale ( $r = .27$ ). This result, plus the fact that the Deference Scale and the ICL Docile-Dependent octant correlate positively (Cairns & Lewis, 1962), would indicate that ICL dominance should correlate negatively with hypnotizability. As seen in Table 1, the correlations obtained in this study were not significant. That there is no consistent relation between dominance and hypnotizability is further supported by Hilgard and Lauer's (1962) CPI data. These authors reported a correlation of  $-.11$  in a sample of 110 males and  $+.23$  in a sample of 106 females.

In a study with 87 female subjects, Schulman and London (1963) found that more susceptible subjects scored significantly lower on the *Pd* scale of the MMPI, indicating that these subjects may be less aggressive, more compliant, and less negative about relations with authority figures than are others. This interpretation of the *Pd* scale results would fit in well with the findings of this study, since the Docile-Dependent, Cooperative-Overconventional, and Responsible-Hypernormal octants of the ICL yielded the highest correlations with hypnosis.

An important next step in further investigating the relationship of interpersonal trust to hypnosis would involve the construction of an item pool which emphasizes the ICL octant items which correlate most highly with hypnosis—JK, LM, and NO. Items such as "appreciative," "cooperative," "considerate," "accepts advice readily," "sociable and

It is difficult to specify which of these differences could account for the differing results; that method variance has as great effects as it does indicates that the "true" correlations between ICL and hypnotic responsiveness are probably quite small. However, since the data of Lang and Lazovik and Schulman and London seem consistent with the present results, the construction of an item pool as described would still seem to be in order.

neighborly," "likes to be taken care of," "agrees with everyone," and "loves everyone" appear in these octants. These items enter into Factor II. It should be pointed out that although items from the Self-effacing-Masochistic quadrant HI enter into Factor II, HI correlations with hypnosis are quite low. Thus, items of the type, "apologetic," "can be obedient," "shy," "meek," or "spineless" should not enter into such a rationally formed item pool. It may be pointed out here, in fact, that the correlations of the sum of the octants JK, LM, and NO with hypnotizability are uniformly higher than the correlations of Factor II which includes the octant HI. For Form A, for example, the correlations in the three samples are .36, .33, and .27 as compared to .31, .32, and .24. The item pool should also contain items from the EPPS Affiliation Scale. It should be pointed out that items of the two octants correlating most highly with hypnosis scores are, on the average, more socially desirable than items of the other octants (Bentler, in press). This characteristic of the items should be taken into account in the construction of an item pool.

Although Form C has been Ås' (1962) preferred scale for correlating hypnotizability and subjective experiences presumably related to hypnosis, the results of this study indicate that Form A may yield as high or higher correlations with personality measures as Form C. The choice of scale for this purpose thus appears to be somewhat arbitrary, though it may well be that finer discriminations may be possible in the future with Form C since it taps a greater variety of hypnotic experiences. It is quite possible, for example, that a person may be able to experience some cognitive alterations without experiencing others; the extent to which these discriminations are important to make would determine the choice of scale.

#### REFERENCES

- ÅS, A. Non-hypnotic experiences related in hypnotizability in male and female college students. *Scand. J. Psychol.*, 1962, 3, 112-121.
- BARBER, T. X. The necessary and sufficient conditions for hypnotic behavior. *Amer. J. clin. Hyp.*, 1960, 3, 31-42.
- BARBER, T. X., & GLASS, L. B. Significant factors in hypnotic behavior. *J. abnorm. soc. Psychol.*, 1962, 64, 222-228.
- BENDIG, A. W. College norms for the concurrent validity of the Pittsburgh revisions of the Maudsley Personality Inventory. *J. psychol. Stud.*, 1959, 11, 12-17.
- BENTLER, P. M. The Interpersonal Check List. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (in press).
- BOUCHER, R. G., & HILGARD, E. R. Volunteer bias in hypnotic experimentation. *Amer. J. clin. Hyp.*, 1962, 5, 49-51.
- CAIRNS, R. B., & LEWIS, M. Dependency and the reinforcement value of a verbal stimulus. *J. consult. Psychol.*, 1962, 26, 1-8.
- CATTELL, R. B., SAUNDERS, D. R., & STICE, G. *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, Illinois: Institute for Personality and Ability Testing, 1953.
- EDWARDS, A. L. *Manual for the Personal Preference Schedule*. New York: Psychological Corporation, 1953.
- ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
- EYSENCK, H. J. *Manual of the Maudsley Personality Inventory*. London: Univ. of London Press, 1959.
- FURNEAUX, W. D., & GIBSON, H. B. The Maudsley Personality Inventory as a predictor of susceptibility to hypnosis. *Int. J. clin. exper. Hyp.*, 1961, 9, 167-177.
- GOUGH, H. G. *California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1957.
- GUILFORD, J. P., & ZIMMERMAN, W. S. *Manual of instructions and interpretations for the Guilford-Zimmerman temperament survey*. Beverly Hills, California: Sheridan Supply Co., 1949.
- HILGARD, E. R. Lawfulness with hypnotic phenomena. In G. H. Estabrooks (Ed.), *Hypnosis: Current problems*. New York: Harper & Row, 1962. Pp. 1-29.
- HILGARD, E. R., & BENTLER, P. M. Predicting hypnotizability from the Maudsley Personality Inventory. *Brit. J. Psychol.*, 1963, 54, 63-69.
- HILGARD, E. R., & LAUER, LILLIAN W. Lack of correlation between the California Psychological Inventory and hypnotic susceptibility. *J. consult. Psychol.*, 1962, 26, 331-335.
- HILGARD, E. R., WEITZENHOFFER, A. M., LANDES, J., & MOORE, ROSEMARIE K. The distribution of susceptibility to hypnosis in a student population: A study using the Stanford Hypnotic Susceptibility Scale. *Psychol. Monogr.*, 1961, 75(8, Whole No. 512).
- LAFORGE, R., & SUCZEK, R. The interpersonal dimension of personality: III. An interpersonal check list. *J. Pers.*, 1955, 24, 94-112.
- LANG, P. J., & LAZOVYK, A. D. Personality and hypnotic susceptibility. *J. consult. Psychol.*, 1962, 26, 317-322.

- LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald Press, 1957.
- SCHULMAN, R. E., & LONDON, P. Hypnotic susceptibility and MMPI profiles. *J. consult. Psychol.*, 1963, 27, 157-160.
- WEBSTER, H., SANFORD, N., & FREEDMAN, M. A new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, 40, 73-84.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale*. Palo Alto, California: Consulting Psychologists Press, Inc., 1959.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto, California: Consulting Psychologists Press, Inc., 1962.
- WEITZENHOFFER, A. M., & WEITZENHOFFER, GENEVA B. Personality and hypnotic susceptibility. *Amer. J. clin. Hyp.*, 1958, 1, 79-82.
- WIGGINS, NANCY. The structure of the Interpersonal Check List. Paper presented at California State Psychological Association, December 1961.
- WHITE, R. W. Two types of hypnotic trance and their personality correlates. *J. Psychol.*, 1937, 3, 279-289.

(Received February 15, 1963)



## RESPONSE SET PARALYSIS: IMPLICATIONS FOR MEASUREMENT AND CONTROL<sup>1</sup>

JEROME SILLER AND ABRAM CHIPMAN

*Engineering Research Division, New York University*

Data on the relationships between 7 response set measures and 42 personality and other variables are reported. 284 college students were administered a 574-item battery, including MMPI-type and Adjective Check List scales, measures of acquiescence, social desirability, and deviance sets, and miscellaneous variables. Analyses of intercorrelations reveal response sets are mostly idiosyncratic to the measure used. These scales, furthermore, account for a small portion of the variance in the personality instruments. An orthogonal factor analysis of the total battery reveals only acquiescence as an independent response set factor. Questions are raised about the utility of burdening test batteries with a mass of response controls.

A series of response set measures was interposed within a large self-report test battery as part of a study of attitudes toward the physically handicapped (Siller, 1962). It is believed that the resulting data have a bearing on two major issues of technical concern in the use of self-report personality questionnaires: (a) What is the relationship between the set to check "true" to items (i.e., acquiescence) and the set to check the socially desirable alternative? (b) In view of the repeated findings of substantial correlations between response sets and personality scales, what is the justification for continued content interpretations of such inventories?

Factor analyses of the MMPI, on which most work has been done, generally find two basic factors to which similar labels are attached (Kassebaum, Couch, & Slater, 1959; Eichman, 1961; Finney, 1961; Welsh, 1956). One is defined in terms of general psychopathology versus ego strength. The second appears to be akin to introversion-extroversion. Edwards interprets the first factor as one of social desirability and the second as acquiescence (Edwards, Diers, &

Walker, 1962; Edwards & Heathers, 1962). Messick and Jackson (1961) believe that "acquiescence, as moderated by item desirability, plays a dominant role in personality inventories like the MMPI." However, there are those who argue that socially desirable responding may relate to the actual absence of psychopathology and anxiety (Crowne & Marlowe, 1960), and that acquiescence may be part of a broader syndrome not unlike introversion-extroversion (Couch & Keniston, 1960).

Specific purposes of this study are to discuss the relationship between seven response set measures on a correlational and factorial level, and to determine their correlations with various personality scales. A final purpose is to determine the factorial consequences of merging adjective check list scales with MMPI-type scales.

### METHOD

A 574-item questionnaire was group administered to 284 students from municipal and community colleges in the New York City area. The sexes were fairly evenly divided and are treated together. The population was almost entirely white and mostly of middle socioeconomic status. Subjects were tested during a regular session of their class. They were informed of the nature of the study, i.e., "to relate the way a person feels about himself to the way he feels about the physically disabled," and a non-defensive test atmosphere was promoted. Subjects responded anonymously.

The selection of a test battery was based on two considerations: the inclusion of a variety of person-

<sup>1</sup> This investigation was supported, in part, by a research grant from the Vocational Rehabilitation Administration, Department of Health, Education, and Welfare, Washington, D. C. We wish to acknowledge our indebtedness to Jacob Cohen for his contribution to the statistical analysis. A brief form of this paper was presented at the 1962 meetings of the American Psychological Association.

ality and stylistic measures, and available classroom time. In selecting response set scales, the criteria of frequency of use and brevity of administration were followed. Thus, the Edwards Social Desirability Scale and an overall agreement score for measuring acquiescence are the ones most frequently referred to and are included. The choice of other response-set measures was importantly determined by the number of new items that would be required in our battery. Inasmuch as the primary focus of the study was on personality factors within the normal range of functioning, the major content instrument employed was the Gough Adjective Check List (ACL). Inclusion of the MMPI was deemed inappropriate because of the absence of marked clinical pathology in the population tested. As markers for MMPI-type variables, however, several scales were retained, such as Welsh's A and R which have been found to load heavily on the two principal factors of the MMPI (Welsh, 1956). Demographic variables and several measures of attitudes to the disabled brought the total number of variables to 49, enumerated as follows:

#### *Demographic and Disability Variables*

1. Age.
2. Sex.
3. Social Distance Scale (Siller, 1962). This is a list of six possible relationships varying in intimacy. The subject must check those to which he would admit disabled persons.
4. Feeling Check List (Siller, 1962). The subject checks those of seven listed reactions to the disabled which apply to his feelings, e.g., repelled, sympathetic, indifferent.
5. Attitudes toward Disabled Persons Scale (ATDP) (Yuker et al., 1960).
6. Disability Familiarity (Siller, 1962). A 3-point scale of the extent of the subject's contact with the handicapped.

#### *Response Set Measures*

1. Marlowe-Crowne Social Desirability (MC-SD) (Crowne & Marlowe, 1960).
2. Edwards Social Desirability (E-SD) (Edwards, 1957).
3. Couch-Keniston Agreement Response Set (CK-ARS) (Couch & Keniston, 1960).
4. Fulkerson Acquiescence (Fulkerson, 1958).
5. Overall Agreement Score for Adjectives (OAS I). The number of "true" responses out of 300 in the Adjective Check List. In our instructions, the subject was to check "true" those adjectives which applied to him, and check "false" those that were not characteristic.
6. Overall Agreement Score for Statements (OAS II). The number of "true" responses out of 197 MMPI-type statements.
7. Grigg & Thorpe Adjective Check List Deviance (Grigg & Thorpe, 1960).

#### *MMPI-type Scales*

1. Welsh Anxiety (A) (Welsh, 1956).
2. Welsh Repression (R) (Welsh, 1956).
3. Barron Ego Strength (Barron, 1953).
4. Zaks-Walters Hostility (Zaks & Walters, 1959).
5. Maslow Security-Insecurity Inventory (Maslow, 1952). First 25 items.
6. Couch-Keniston Low F (Couch and Keniston, 1960).

#### *Adjective Check List Scales*

1. Zuckerman Affect Adjective Check List (AACL) (Zuckerman, 1960).
2. Berdie Femininity (Berdie, 1959). Scales 3 through 16 are from Gough, 1955.<sup>2</sup>
3. Rigidity.
4. Positive Character Integration.
5. Self-Insight.
6. Self-Acceptance.
7. Flexibility.
8. Role Playing.
9. Social Poise and Presence.
10. Potential Success.
11. Originality.
12. Self-Criticism.
13. Good Judgment.
14. Personal Adjustment.
15. Likeability.
16. Responsibility.

Scales 17 through 30 are from Heilbrun, 1959.

17. Aggression.
18. Heterosexuality.
19. Dominance.
20. Exhibition.
21. Achievement.
22. Deference.
23. Nurturance.
24. Change.
25. Endurance.
26. Abasement.
27. Intraception.
28. Succorance.
29. Autonomy.
30. Affiliation.

The instructions here employed for the ACL differed from those usually given. Subjects were instructed not merely to check only those adjectives which applied to them but to check every item as being true or false of them. The present method permits both a direct acquiescence index, based on the number of true's on the ACL, and, in our opinion, a more effective control of acquiescence within each scale than the standard score conversion for acquiescence recommended by Heilbrun (1959). Our major criticism of Heilbrun's standard con-

<sup>2</sup> A revised version of the ACL with several different scales and changed scoring is soon to appear. The scales employed in this study consist of those mentioned in Gough (1955) and some sent by him to us since then.

TABLE 1  
INTERCORRELATIONS OF RESPONSE SET WITH PERSONALITY AND OTHER VARIABLES  
(*N* = 284)

	Edw.- SD	MC- SD	Fulk. Acq.	CK- ARS	OAS I	OAS II	ACL Dev.
Edwards SD							
Marlowe-Crowne SD	38						
Fulkerson Acquiescence	-43	-31					
Couch-Keniston ARS	-39	-14	46				
Overall Agreement Score I	-25	-08	39	43			
Overall Agreement Score II	-75	-27	67	57	42		
Grigg-Thorpe ACL Deviance	-54	-32	27	32	30	44	
Age	16	03	-06	-20	-03	-19	-08
Sex	-06	03	-12	-07	-08	-01	-03
Social Distance Scale	06	00	-04	11	04	-02	03
Feeling Check List	03	17	-01	-05	04	-03	00
ATDP	15	09	-07	-16	-10	-13	-12
Disability Familiarity	00	-03	08	06	10	07	-02
Barron Ego Strength	67	15	-35	-26	-20	-57	-33
Welsh Anxiety	-88	-45	50	43	22	79	51
Welsh Repression	14	05	-48	-35	-40	-53	-09
Zaks-Walters Hostility	63	-29	55	53	30	73	41
Maslow Security-Insecurity	70	37	-23	-22	-12	-49	-48
Zuckerman AACL	-67	-37	27	23	26	48	60
Couch-Keniston Low F	-22	-14	12	13	15	19	18
Berdie Femininity	-18	-08	08	13	29	20	11
ACL Rigidity	00	08	-12	-20	-12	-09	-10
ACL Positive Character							
Integration	41	32	-30	-42	-47	-44	-56
ACL Self Insight	43	25	-23	-22	-19	-38	-59
ACL Self Acceptance	42	31	-06	-06	22	-20	-61
ACL Flexibility	09	03	02	06	13	00	-09
ACL Role Playing	25	22	-31	-29	-48	-33	-31
ACL Social Poise and Presence	50	34	-26	-26	-31	-40	-55
ACL Potential Success	42	25	-36	-44	-54	-47	-44
ACL Originality	08	-04	-21	-36	-41	-29	-07
ACL Self Criticism	-59	-37	40	43	60	56	75
ACL Good Judgment	38	25	-27	-39	-32	-38	-53
ACL Personal Adjustment	51	37	-25	-28	-22	-39	-74
ACL Likeability	49	40	-23	-23	-15	-36	-66
ACL Responsibility	36	29	-30	-38	-27	-37	-58
ACL Aggression	-26	-33	26	19	37	25	51
ACL Heterosexuality	21	10	07	10	22	-04	-30
ACL Dominance	39	15	00	-05	11	-25	-37
ACL Exhibition	10	-04	15	15	29	04	05
ACL Achievement	45	27	-05	-18	16	-30	-57
ACL Deference	-02	14	-23	-12	-35	-10	-22
ACL Nurturance	44	39	-29	-25	-38	-35	-73
ACL Change	-08	-10	21	14	14	16	-02
ACL Endurance	50	37	-26	-33	-10	-43	-65
ACL Abasement	-40	-13	01	10	-12	23	25
ACL Intraception	50	32	-23	-27	-12	-38	-71
ACL Succorance	-58	-31	25	26	30	43	58
ACL Autonomy	-06	-16	24	13	44	18	29
ACL Affiliation	34	33	-17	-07	-08	-18	-64



TABLE 2  
QUARTIMAX-ROTATED CENTROID FACTOR LOADINGS  
( $N = 284$ )

	I	II	III	IV	$h^2$
Adjective Check List: Likability	86	-09	13	10	78
Grigg & Thorpe: ACL Deviance	-84	04	-02	-03	70
ACL: Personal Adjustment	83	-10	10	12	73
ACL: Self Criticism	-81	15	32	-13	79
ACL: Nurturance	80	-27	01	39	87
ACL: Intraception	79	-07	15	-04	65
ACL: Social Poise and Presence	77	-08	-12	03	61
ACL: Endurance	76	-02	07	-40	74
Zuckerman: Affect Adjective Check List	-75	-17	19	-05	63
ACL: Self Acceptance	73	17	42	-07	75
ACL: Succorance	-71	-28	18	01	61
ACL: Self Insight	70	-08	02	-18	53
Edwards: Social Desirability	70	32	-39	-18	77
ACL: Affiliation	70	-07	21	45	74
Maslow: Security-Insecurity Inventory	68	39	-17	01	64
Welsh: Anxiety	-68	-38	44	19	83
ACL: Responsibility	66	-42	00	-27	67
ACL: Achievement	62	40	23	-26	66
ACL: Potential Success	53	-23	-40	-14	51
Zaks-Walters: Hostility	-51	-03	43	12	47
Marlowe-Crowne: Social Desirability	47	01	-12	04	24
Barron: Ego Strength	41	24	-37	-13	43
ACL: Role Playing	37	-24	-34	23	36
Couch-Keniston: Low I'	-20	-01	11	08	06
ACL: Abasement	-29	-80	05	04	73
ACL: Exhibition	-05	77	08	26	67
ACL: Autonomy	-32	75	11	-11	69
ACL: Aggression	-59	62	06	-13	76
ACL: Dominance	43	56	10	-11	52
ACL: Change	-11	50	15	40	44
ACL: Heterosexuality	32	40	28	39	49
Agreement Score: ACL-type items	-24	33	67	-14	63
Agreement Score: MMPI-type items	-55	-09	65	19	76
Fulkerson: Acquiescence	-35	13	56	05	46
Welsh: Repression	08	-30	-52	-01	36
ACL: Rigidity	12	-49	-04	-51	52
Berdie: Femininity (ACL)	-14	-01	25	39	24
Attitudes Toward Disabled Persons	17	11	-12	30	14
Feeling Check List	08	19	-03	25	11
Social Distance Scale	-01	06	-01	21	05
$\Sigma a^2$	12.51	4.64	3.18	2.01	22.34
Percentage of Variability	55.99	20.76	14.23	9.01	100.00

version technique is the fact that only four categories of number of words checked are provided, the highest of which covers the extremely heterogeneous score range of 122-300. Raw scores were used throughout our study, and our data are therefore not directly comparable to what would be given by Heilbrun's method.

A matrix of Pearson product-moment coefficients of correlation was computed and factored by the

centroid method, using the highest coefficient per column as the communality estimate, without iteration. The factors were rotated by the quartimax analytic rotational procedure (Harmon, 1960). All computations were performed on the IBM 650 electronic computer. The capacity of the computing program available for centroid extraction is 40 variables. Nine of the 49 variables judged least important were omitted from the factor analysis. The

Couch-Keniston ARS was one of those dropped because a similar, larger scale (OAS II) was in the matrix and because the pattern of correlations for ARS was virtually identical to that of the Fulkerson scale. Although separate analyses were not run for each sex, the original matrix included sex as a variable. The few sex differences found in the inter-correlations were similar to those found for the Berdie Femininity Scale which, as a marker variable, is retained in the  $40 \times 40$  factor analysis.

Materials on disability are included in the tables but not discussed here as they form a unique factor extraneous to this study.

## RESULTS AND DISCUSSION

We are concerned in this report only with the zero-order correlations and the first three rotated factors. These factors account for 31, 12, and 8% of the total variance, and 56, 21, and 14% of the common variance, respectively. Table 1 contains the zero-order correlations between the response set and other measures. Table 2 presents the results of the factor analysis.

In support of similar findings by McGee (1962) and Foster (1961), only modest correlations between acquiescence measures were found. The highest correlations were between the Fulkerson Scale and the OAS II (+.67) and between the CK-ARS and the OAS II (+.57). All other correlations were between .39 and .46. The two measures of SD show a low but significant relationship ( $r = +.38$ ). The most appreciable correlation is between the Edwards SD scale and the OAS II (-.75). The last correlation is understandable when one notes that 30 of the 39 items on the Edwards SD scale are keyed false and that the OAS II is heavily loaded on pathology acceptance. By consistently denying pathology, one would receive a high Edwards score and a low OAS II score. To a considerable extent, then, response sets are mostly idiosyncratic to the measure used.

While the expected negative relationships between acquiescence and social desirability occur in our data, it is essential to temper our interpretation of this with considerations of the relative worth of the different indices of each response set. For the delineation of SD as a source of error, Marlowe and Crowne felt their own SD scale to improve upon major failings of the Edwards. Their

argument is that Edwards' scale requires the subject to deny psychopathology at the same time he is affirming socially approved behaviors. It is most parsimonious for non-psychiatric populations to assume that actual absence of pathology may be the causal basis for SD scores, even though SD ratings show high correlations. Marlowe and Crowne, in their scale, avoid this confounding effect by requiring the subject to base his avowals of SD on behaviors which are not merely "normal" or "healthy" but up to a standard of perfection which is generally improbable in a mere mortal. The rejoinder of Edwards, Diers, and Walker (1962) that the MC-SD likely is a lie scale and therefore not comparable with the Edwards as an SD measure does not alter the situation, since if the subject is *not* lying, his giving of the SD response does not constitute a source of contamination of the content scale. Hence, we regard the M-C as the preferable SD indicator because of this freedom from the confounding effect of psychopathology and because of its additional merit of evenly balancing items keyed true (18) and items keyed false (15) whereas Edwards' scale has a predominance of keyed false items. As to the relationships with acquiescence measures, the average correlation of the M-C is  $-.20$  as opposed to a value of  $-.46$  for the Edwards.

A similar distinction can be made between the different measures of acquiescence. Agreement scores based on MMPI-type items are subject to the criticism that answering true consistently implies acceptance of pathological and/or socially undesirable item content. Hence it is not surprising that an MMPI-type agreement score (OAS II) correlates  $-.75$  with the Edwards scale and  $-.27$  with the Marlowe-Crowne. The ACL-derived agreement score (OAS I) is less contaminated in this manner inasmuch as the ACL focuses on traits within the normal domain of behavior and contains as many apparently negative traits as positive ones.<sup>3</sup> The correlations be-

<sup>3</sup> Objection may be made to this in the light of Heilbrun's finding (1959) of a rho correlation of .98 between frequency of adjective endorsement and judged social desirability of the adjectives. Heilbrun

tween this scale and SD are  $-.25$  for Edwards SD scale and  $-.08$  for Marlowe-Crowne. The strong impression of these findings is that the shared variance of acquiescence and social desirability approaches zero, when the purest measure of both is used.

Turning from zero-order correlations to factorial structure, it is found that the acquiescence measures load principally on the third factor and entirely account for it. The correlations between proportions of items in a scale keyed true and that scale's loading on a factor are:  $+.90$  for Factor III,  $+.10$  for Factor II, and  $-.39$  for Factor I. At this point we conclude, unlike Messick and Jackson (1961), that acquiescence is relatively unimportant for Factors I and II. Evidence in the next section utilizing content scales reinforces this position. SD has its highest impact on the first factor, with loadings of  $+.70$  and  $+.47$  for the Edwards and Marlowe-Crowne, respectively. The loadings on all the factors account for only 24% of the total variance of the Marlowe-Crowne (the better SD measure). Hence there is a strong implication that SD as such is a minor contributor to our battery.

#### *Relationship between Content Scales and Response Sets*

The average zero-order correlation with personality measures for each response set measure was: Edwards SD  $+.39$ ; Marlowe-Crowne SD  $+.24$ ; Couch-Keniston Acq.  $+.24$ ; Fulkerson Acq.  $+.23$ ; OAS I  $+.27$ ; OAS II  $+.34$ ; Deviation Score  $+.43$ .

It has already been indicated why we believe that Factor I is not interpretable in terms of social desirability or acquiescence. Before considering a content interpretation, the  $-.84$  loading of the ACL Deviance Scale must be evaluated. Since ample evidence exists to demonstrate that the Deviation Hypothesis holds even for noncontent ma-

terial, one might interpret Factor I as a tendency to deviate. But why do people deviate? It is not feasible to deal with other studies of the Deviation Hypothesis; however, an answer for the specific measure used here is possible. The ACL Deviance Scale was validated on college students requiring mental hygiene help versus those not requiring such help. Therefore these groups presumably had greater and lesser degrees of ego strength, adaptability, anxiety, etc. This indeed is what is found on Factor I. It is concluded then that a nonresponse-set interpretation of this factor in terms of Ego Strength-Ego Weakness is possible.

Having identified Factor I as Ego Strength-Ego Weakness, Factor III as Acquiescence, and Factor IV as Acceptance of the Physically Disabled, we are left only with the interpretation of Factor II. Factor II loads most heavily on variables relating to ascendance, i.e., Dominance, Exhibition, Aggression, and negatively on Abasement. However, a high positive loading on Autonomy complicates interpreting the Factor as introversion-extroversion. Recently Factor II in self-report questionnaire studies has been interpreted instead as Acquiescence. The fact that we have only a few of the MMPI scales undoubtedly has contributed to these discrepant findings. However, the Welsh A and R scales, generally considered anchor points for MMPI Factors I and II, were included. Inclusion of the ACL seems to have provided a new perspective on these issues. We find that acquiescence splits off, probably from Factor II, and becomes a separate third factor.

Our review of previous literature in this area had led us to certain conclusions. The main ones were that if response sets were ignored, as almost all factor analyses of the MMPI have done, the principal issues were the interpretation of Factor III and the effect of oblique rotations. If a study included response sets, the impulse was to think only in noncontent terms. The present study once again indicates the dangers of overgeneralizing to the whole realm of personality structure from a relatively restricted range of test data; i.e., factors that could be so changed by the addition of differently styled scales could

questions the predictive meaning of this correlation and in a subsequent study (Heilbrun, 1962) demonstrates that an ACL Achievement scale, not corrected for SD, actually has better predictive value than an EPPS Achievement scale that is corrected for SD. This is entirely in accord with the simple notion that most people act in terms of their values and need for social approval.



hardly be representative of basic personality dimensions.

It has not been our intention to deny that measures of response sets can account for some of the variance of self-report inventories. It certainly is apparent that new scales should be balanced for true and false in their keying, and that social desirability be given particular attention. Nevertheless, it is our contention that content does remain the major component, and that the first two factors of self-report batteries are not basically interpretable in terms of social desirability and acquiescence. We are inclined to feel that the overvaluation of these response set factors has caused undue professional effort to be diverted from the real task of developing rich but rigorous measures of personality functioning.

#### REFERENCES

- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- BERDIE, R. F. A femininity adjective check list. *J. appl. Psychol.*, 1959, 43, 327-333.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, 56, 81-105.
- CATTELL, R. B. Foundations of personality measurement theory in multivariate experiment. In I. G. Sarason (Ed.), *Contemporary research in personality*. Princeton: D. Van Nostrand, 1962. Pp. 99-111.
- COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, 60, 151-174.
- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, 24, 349-354.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden Press, 1957.
- EDWARDS, A. L., DIERS, CAROL, & WALKER, J. Response sets and factor loadings on sixty-one personality scales. *J. appl. Psychol.*, 1962, 46, 220-225.
- EDWARDS, A. L., & HEATHERS, LOUISE B. The first factor of the MMPI: Social desirability or ego strength? *J. consult. Psychol.*, 1962, 26, 99-100.
- EICHMAN, W. J. Replicated factors on the MMPI with female NP patients. *J. consult. Psychol.*, 1961, 25, 55-60.
- FINNEY, J. C. The MMPI as a measure of character structure as revealed by factor analysis. *J. consult. Psychol.*, 1961, 25, 327-336.
- FOSTER, R. J. Acquiescent response set as a measure of acquiescence. *J. abnorm. soc. Psychol.*, 1961, 63, 155-160.
- FULKERSON, S. C. An acquiescence key for the MMPI. *Sch. Aviat. Med.*, 1958, 1-11.
- GOUGH, H. G. *Reference handbook for the Gough Adjective Check-List*. Berkeley: Univer. Calif., Institute Personal Assessment & Research, 1955.
- GRIGG, A. E., & THORPE, J. S. Deviant response in college adjustment clients: A test of Berg's deviation hypothesis. *J. consult. Psychol.*, 1960, 24, 92-94.
- HARMON, H. *Modern factor analysis*. Chicago: Univer. Chicago Press, 1960.
- HEILBRUN, A. B., JR. Validation of a need scaling technique for the Adjective Check List. *J. consult. Psychol.*, 1959, 23, 347-351.
- HEILBRUN, A. B., JR. Social desirability and the relative validities of achievement scales. *J. consult. Psychol.*, 1962, 26, 383-386.
- KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, 23, 226-236.
- MASLOW, A. H. *Manual for the security-insecurity inventory*. Stanford, California: Stanford Univer. Press, 1952.
- MCGEE, R. K. The relationship between response styles and personality variables: I. The measurement of response acquiescence. *J. abnorm. soc. Psychol.*, 1962, 64, 229-233.
- MESSICK, S. J., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychol. Bull.*, 1961, 58, 299-304.
- SILLER, J. Personality determinants of reaction to the physically handicapped. *Amer. Psychologist*, 1962, 17, 338. (Abstract)
- WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- YUKER, H. E., BLOCK, J. R., & CAMPBELL, W. J. *A scale to measure attitudes toward disabled persons*. Albertson, New York: Human Resources Foundation, Division of Abilities, Inc., 1960.
- ZAKS, M., & WALTERS, R. H. First steps in the construction of a scale for the measurement of aggression. *J. Psychol.*, 1959, 47, 199-208.
- ZUCKERMAN, M. The development of an affect adjective check list for the measurement of anxiety. *J. consult. Psychol.*, 1960, 24, 457-462.

(Received February 25, 1963)

## COMPARATIVE ANALYSIS OF PSYCHOTIC DEPRESSIVES WITH MATCHED NORMALS ON SOME UNTIMED VERBAL INTELLIGENCE TESTS<sup>1</sup>

SAMUEL GRANICK,

*Philadelphia Psychiatric Center*

50 psychotic depressives, matched to a group of 50 normals for age, sex, race, education, religion, and nativity, are compared for performance on WAIS Information and Similarities, and on Thorndike-Gallup Vocabulary. Intergroup differences are not significant, but the normals tend to score higher on all tests. Pattern of performance is similar for both groups. Weighted information scores are significantly higher than Similarities scores, at the .01 level, for both the normals and depressives. A precipitate drop in quality of performance on the Similarities is found in relation to increasing age within the 2 groups. The age factor is, thus, found to be more potent than depression in producing psychometric deficit, particularly in the area of abstraction and flexibility of thought processes.

### METHOD

#### *Subjects*

The psychotic depressive subjects were selected from a pool of mental hospital patients carefully screened and diagnosed by two experienced psychiatrists. Normals, matched with the patient group for sex, age, religion, race, and nativity, were chosen from a larger sample which had been recruited through Golden Age Clubs, other types of adult social organizations, and two employment services. Information elicited from these subjects indicated that they had no history of mental disorders or hospitalizations for emotional conditions. All subjects, furthermore, were white and judged to be of middle class socioeconomic status. The other background variables are summarized in Table 1. Equivalence of intellectual capacity was assumed for the two groups because of the close matching achieved, particularly with respect to age and education.

#### *Procedure*

The WAIS Information and Similarities and the Thorndike-Gallup Vocabulary Tests were administered to the subjects individually by the same female examiner and in the same order as part of a much larger battery of cognitive, perceptual, psychomotor, and affective reaction tests. Each patient was examined shortly after admission to the hospital, following diagnostic evaluation by the two psychiatrists associated with the research project. Medication was withheld for at least a 24-hour period so as to eliminate or reduce any possible effects which drugs might have on test performance. The normal subjects came to the hospital research building individually, where they were tested under conditions which were fairly identical to those set up for the patient group. It was explained to them that they were participating in a research program.

Availability of systematically collected data on the intellectual functioning of psychotic depressives is sparse. Summaries of research in this area by Hunt and Cofer (1944) and more recently by Payne (1961) indicate that results can be regarded only as tentative and inconclusive since they are based on studies with very small and limited samples. Generally, however, the impression is that cognitive and thinking processes of depressives are not deteriorated or regressed or remarkably different from normals when speed of performance is not a factor and testing conditions are supportive. The present report aims to investigate this impression in terms of the following questions: (a) Do psychotic depressives perform as well as normals on such untimed, verbal intelligence test items as the Wechsler Adult Intelligence Scale (WAIS) Information and Similarities Tests, and the Thorndike-Gallup Vocabulary Test? (b) How do the two groups compare in their interest pattern of functioning? (c) What are the relationships among test performances, age, and education for each group of subjects?

<sup>1</sup> Based on a research project entitled "Vigilance, Sedation Threshold, Tofranil and Depression," supported by Grant MY 3674 of The National Institute of Mental Health.

Thanks are due to Alfred S. Friedman for cooperation and help in the research and preparation of this report. Ethel Shapiro served as the psychometrician who tested all the subjects.



TABLE 1  
BACKGROUND VARIABLES

<i>N</i>	Psychotic depressives ( <i>N</i> = 50)	Normals ( <i>N</i> = 50)
Male	21	21
Female	29	29
Mean age	57.26	57.36
<i>SD</i>	7.81	7.15
Mean education (grade)	8.47	8.82
<i>SD</i>	3.54	3.17
Native born	40	40
Foreign born	10	10
Jewish	34	34
Christian	16	16

A nominal compensation for their time was offered, which most accepted either for their own use or for a contribution to a charitable program. The patient group understood that the testing for them was part of the hospital routine, but generally there seemed to be an awareness that research was involved.

### RESULTS

In Table 2, the comparison between the two groups on their raw test scores shows the normals consistently performing better than the psychotic depressives. The differences, however, are not striking and in only one instance is statistical significance approached. (Information is a shade below the .05 level.) Noteworthy also is the closeness of the means and *SD*s on Vocabulary. This confirms the assumption mentioned above of similarity of intelligence between the two groups, based on

TABLE 2

COMPARISON OF RAW TEST SCORES BETWEEN  
PSYCHOTIC DEPRESSIVES, AND NORMAL SUBJECTS  
(*N* = 50)

Tests	Psychotic depressives		Normals		<i>t</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Information	14.28	5.0	16.20	5.14	1.94
Similarities	8.64	5.88	10.18	5.97	1.30
Thorndike-Gallup	10.86	3.87	11.28	4.03	0.53

the matching procedure, since the Thorndike-Gallup Vocabulary Test was designed as an intelligence measure (Miner, 1957).

A further examination of the data, on the other hand, along with the intercorrelations in Table 3, impresses one with the relative similarity in the pattern of performance between the normals and the depressives. Characteristically, however, the normals seem more consistent and approximate more closely the relationships among the tests reported by Wechsler (1958), and Doppelt and Wallace (1955) for the age range covered by the subjects. The rather substantial correlations between education and each of the three tests are noteworthy. Here, too, the normals perform more like Wechsler's national standardization sample for the WAIS than the depressives, but the latter show the same basic pattern.

The low negative, statistically insignificant correlation of age and education is the same for both groups. In this, there is a difference from the *r* of  $-.29$  reported by Birren and Morrison (1961) in their analysis of 933 native white subjects in the age range of 25 to 64 years from the WAIS standardization sample. Accounting for it, probably, is the relatively narrower age range of the subjects of the present study, which did not permit the well-known historical trend toward increasing educational opportunities to show itself.

Intragroup analysis of the test scores shows a striking difference between performance on the Information and Similarities Tests. The average weighted scores for the tests, as shown in Table 4, are different to a very significant degree. Reflected herein is a large decline in quality of functioning on Similarities by the population of which the study samples are representative. The correlations in Table 3 suggest that the age factor is most significant in this decline. The three tests show some decline with age for both groups, but it is quite small and not statistically significant for Vocabulary and for Information in the case of the normal group. The *r* of  $-.30$  between Information age for the depressives is significant at the .05 level. But for Similarities, a precipitate decline with age is found in both groups, the *r* for the



TABLE 3  
INTERCORRELATIONS AMONG TEST SCORES, AGE, AND EDUCATION

	Depressives				Normals			
	Information	Similarities	Vocation	Education	Information	Similarities	Vocation	Education
Age	-.30	-.46	-.22	-.16	-.19	-.33	-.02	-.15
Education	.58	.44	.39		.61	.63	.68	
Vocation	.61	.34			.65	.64		
Similarities	.56				.58			

depressed subjects being significant at better than the .01 level and for the normals at about the .03 level.

These results, particularly for the normal group, led to a more detailed analysis of the quality of the subjects' responses to the individual items of the Similarities test. Table 5 shows the percentages correct on each item for both groups in comparison with the Wechsler data.

There are, of course, many more significant differences between the depressed sample and the Wechsler norms than between the normals and Wechsler's group. The remarkable element, however, is the exceedingly large incidence of failures on the relatively easy items of the test by the subjects of both groups. Also of interest is the fact that successes on the first three items for both groups included over 40% of scores of "one," reflecting rather concrete, relatively immature performance.

#### DISCUSSION

The data are consistent with the noted impression above, derived from summaries of the research literature, that psychotic de-

pressives are not significantly impaired in cognitive functions of the type that the Information, Similarities, and Vocabulary Tests represent. Agreement is also found with Friedman's (in press) forthcoming report, based on research of which the present study is a part, containing an extensive demonstration of the basic intellectual intactness of hospitalized depressed subjects.

Yet it cannot be ignored that the normals scored consistently higher on all three of the tests. Friedman also noted such a trend, reporting the normals significantly superior on 11 of the 82 scores analyzed, and the depressives generally tending toward erratic performance. Depression, therefore, appears to take its toll, but when tested under benign

TABLE 5  
COMPARISON OF PERCENTAGE OF CORRECT RESPONSES ON SIMILARITIES TEST

Item	Depressives	Normals	Wechsler sample (N=1500)
	% Correct	% Correct	% Correct
1	70*	78*	93
2	88	88	90
3	62*	62*	90
4	66*	74**	86
5	56*	66	73
6	42*	50*	69
7	36*	48	56
8	50	46	55
9	22*	40	46
10	22**	44	38
11	16	10	21
12	10**	22	25
13	10	14	18

\* Significantly different from Wechsler data at .01 level.  
\*\* Significantly different from Wechsler data at .05 level.

TABLE 4

COMPARISON OF WEIGHTED SCORES FOR INFORMATION AND SIMILARITIES OF PSYCHOTIC DEPRESSIVES AND NORMALS (N = 50)

	Information		Similarities		t
	M	SD	M	SD	
Depressives	9.32	2.57	7.16	3.31	3.65*
Normals	10.24	2.60	7.96	3.50	3.62*

\*  $p < .01$ .

conditions (as is the case in most clinically oriented studies), the patient can perform at close to normal levels. The present study, furthermore, suggests that as the complexity of mental operations increases, the depressed subject finds it more difficult to use his capacities. Almost identical means and SDs were achieved by the two groups on the Vocabulary Test, a situation in which only recognition is involved in a multiple choice of synonyms. But the far more difficult and complex operations of recall and abstraction are involved respectively on Information and Similarities, requiring the exercise of sustained effort and some flexibility of mental functioning. The relatively inferior quality of performance on these tests by the depressives as compared to the normals leads one to anticipate that under stressful circumstances, particularly when the facilitating support of the examiner is absent, the depressed subjects would show a more marked and significant intellectual deficit.

Another feature deserving attention is the poor quality of performance on Similarities by the subjects in both groups. From the data available, the age factor seems to be most prominently associated with this phenomenon, suggesting a sharp decline of functioning as the individual moves from the fifth and sixth to the seventh and eighth decades of life. The educational factor is not significant in the decline, despite its high correlations with the test scores, since the relationship between age and education is small for the samples. Thus, when education is partialled out, the  $r$ 's reduce only to  $-.44$  and  $-.31$  for the depressed and normal groups, respectively.

A marked decline with age on Similarities was not found in the original standardization group, which included individuals up to age 64. But Doppelt and Wallace (1955) observed it in their supplementary standardization group, which consisted of subjects from age 60 to over 75. No rationale was offered for the decline. They did comment, however, that other verbal tests, such as Vocabulary, Information, and Comprehension, are useful to the elderly in communica-

tion, but that Similarities "constitutes an artificial situation by the standards of everyday life." Motivation might thus be an important factor. If so, it is doubtful whether it is of great significance, since no direct evidence has been presented showing that the aged extend themselves less conscientiously on this test than on the others. The very erratic quality of performance shown by the subjects of the present study, on the other hand, particularly in their handling of some of the relatively easy items of the test, suggests that there is probably a factor of deterioration involved. One can only speculate, however, at this point, as to whether this is psychological or organic or perhaps both.

Tapped by Similarities are such mental processes as abstraction, reasoning, and imagination. Flexibility must be exercised by the subject. These are mental functions which have been variously reported (Jones, 1959; Schaie, 1958) to decline significantly with advancing age; but such decline is also associated clinically with cerebral defects. Additionally, consideration must be given to the fact that later maturity is a period when brain deterioration is highly prevalent. Accordingly, one is tempted to hypothesize that the results of the study are largely a reflection of organic deterioration with secondary loss of psychological integration. Further investigation of this area would seem to be both important and fruitful.

#### REFERENCES

- BIRREN, J. E., & MORRISON, D. F. Analysis of the WAIS subtests in relation to age and education. *J. Gerontol.*, 1961, 16, 363-369.
- DOPPELT, J. E., & WALLACE, W. L. Standardization of the Wechsler Adult Intelligence Scale for older persons. *J. abnorm. soc. Psychol.*, 1955, 51, 312-330.
- FRIEDMAN, A. S. Minimal effects of severe depression on cognitive functioning. *J. abnorm. soc. Psychol.*, in press.
- HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and behavior disorders*. New York: Ronald Press, 1944.
- JONES, H. Intelligence and problem solving. In J. E. Birren (Ed.), *Handbook of aging and the individual*. Chicago: Univ. Chicago Press, 1959.

- MINER, J. B. *Intelligence in the United States*. New York: Springer, 1957.
- PAYNE, R. N. Cognitive abnormalities. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961.
- SCHALE, K. W. Rigidity-flexibility and intelligence: A cross-sectional study of the adult life span from 20 to 70 years. *Psychol. Monogr.*, 1958, 72(9, Whole No. 462).
- WECHSLER, D. *The measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins, 1958.

(Received December 31, 1962)

#### ERRATUM

In the article by Vogel and Lauterbach in the *Journal of Consulting Psychology* (1963, 27, 236-242), the authors note that "Schooler, 1961" should be deleted as a reference for the statement on page 236, line 38 and again on page 238, line 76. Schooler's finding that schizophrenics are born in the second half of the birth order was true for all families, independent of family size, and was not limited to families of five or more sibs, as stated in the article.



## NOTES AND COMMENTS

### USE OF THE SHIPLEY-HARTFORD CONCEPTUAL QUOTIENT AS A MEASURE OF INTELLECTUAL IMPAIRMENT<sup>1</sup>

PETER M. LEWINSOHN

*Larue D. Carter Memorial Hospital*

A study to evaluate the use of the Shipley-Hartford (SH) Conceptual Quotient (CQ) as a measure of intellectual impairment. 4 groups of 30 Ss, matched for age and vocabulary, were given the SH. Chronic (C) and acute (A) schizophrenics (schiz), nonschizophrenic patients, and normal Ss were used. Residual abstraction scores (RA) were computed for each S. Results were as follows: (a) CQs and RA scores of C and A schiz were lower than those of normal Ss at the .01 and .05 level, respectively; (b) A correlation of  $-.25$ , significant at .01 level, was obtained between RA and a measure of conceptual disorganization for a group of 137 mental patients. In contrast to the CQ, the RA is independent of age and vocabulary.

The use of the Shipley-Hartford (SH) Conceptual Quotient (CQ) as a measure of intellectual impairment is based upon the assumption that where there is intellectual impairment, vocabulary is less affected than is the capacity for abstract thinking and that in such cases there will be a discrepancy between vocabulary level and ability to handle abstract problems. While direct evidence for this assumption has never been provided, Shipley (1941) was able to demonstrate rather consistent differences in CQ between a normal standardization group and various groups of psychiatric patients who were presumed to show varying degrees of intellectual impairment. In Shipley's study, psychopathic personalities, psychoneurotics, and alcoholics did not differ from the standardization group, while the lowest CQs were obtained by organics, followed by the functional psychotics.

The failure of Magaret and Simpson (1948) to find a correlation between the CQ and psychiatrists' ratings of "deterioration" casts some doubt on the assumption that the discrepancy between vocabulary and abstraction levels is proportionate to intellectual impairment. (It should be pointed out, however, that the psychiatrists' ratings were done 6 months after the test was taken and no evidence for the

reliability of the ratings was presented.) Perhaps potentially more damaging to this assumption is the existence of a correlation of  $-.43$  between age and CQ (Fecher, 1946, Garfield & Fey, 1948) and of a correlation of  $.30$  between vocabulary and CQ (Lewinsohn & Nichols, 1961).<sup>2</sup> Thus the CQ normally declines with age and decreasing vocabulary level. Since Shipley's (1941) original groups were not matched for age or vocabulary level (the normal standardization group was young and it is likely that some of the organic and psychotic groups were older and had a lower vocabulary level than some of the less disturbed groups), it is at least possible that some of the differences in CQ between Shipley's groups were due to differences in age and vocabulary level between the groups.

The present study was to determine the discriminative value of the CQ when age and vocabulary were controlled. The null hypotheses tested were:

1. The CQs of chronic and acute schizophrenics do not differ from those of appropriate control subjects.
2. The discrepancy between performances on the abstraction and vocabulary subtests of the S-H is unrelated to conceptual disorganization.

#### METHOD

First, four groups of 30 subjects, approximately matched for age and vocabulary, were selected.

<sup>2</sup> P. M. Lewinsohn and R. C. Nichols, unpublished results, 1961.

<sup>1</sup> This study was partially supported by a grant from the National Institute of Mental Health (M6029).

The terms impairment and conceptual disorganization are used to denote loss of efficiency without implying irreversibility.

TABLE 1  
MEAN AGE AND VOCABULARY SCORES FOR THE GROUPS

Group	N	Age		Vocabulary	
		M	SD	M	SD
Acute schizophrenics					
Male	15	34.0	8.9	23.9	6.5
Female	15	35.7	4.8	25.1	4.2
Nonschizophrenic psychiatric					
Male	15	34.0	8.2	26.1	6.5
Female	15	35.7	5.3	24.3	4.5
Chronic schizophrenics					
Male	15	34.1	8.2	24.1	5.5
Female	15	36.7	4.8	25.0	4.1
Normal control					
Male	15	33.3	9.2	24.7	5.7
Female	15	35.1	5.4	25.9	4.6

Group 1 consisted of newly admitted acute schizophrenics who were carefully selected on the basis of diagnosis and symptoms. Group 2 was a heterogeneous group of psychiatric patients from the same hospital who had received a diagnosis other than schizophrenia. Patients manifesting schizophrenic symptoms were excluded from this group. Group 3 consisted of chronic schizophrenics from a state hospital<sup>3</sup> who had been ill for at least 6 years, had been continuously hospitalized for at least 3 years, and whose history included evidence of disturbed thinking and bizarre behavior. No patient suspected of being brain damaged was included in this study. Group 4 was selected from a large and heterogeneous group of students, general medical patients, and hospital attendants. Subjects in the four groups were individually paired for sex, age, and S-H vocabulary score. The results of the matching are shown in Table 1.

The S-H was administered individually to the chronic schizophrenics and in small groups to the other subjects using standard directions for administration (Shipley, 1940).

Next, the age and S-H scores of 180 consecutive admissions to a psychiatric hospital were used to construct a multiple regression equation to predict S-H abstraction scores from age and vocabulary.<sup>4</sup> An estimated abstraction score was then computed for each subject and subtracted from his actual abstraction score. This residual abstraction score (RA) thus represents the amount by which the

subject's abstraction score is above or below the score which would be predicted for him on the basis of age and vocabulary score. RA scores were also computed for subjects in the four matched groups.

Finally, ratings on the Multidimensional Scale for Rating Psychiatric Patients (Lorr, 1953) for 137 of the 180 consecutive admittees were used to calculate the correlation between RA scores and Lorr's Factor K (conceptual disorganization). These ratings were made on the basis of an independently conducted interview shortly after admission to the hospital.

## RESULTS AND DISCUSSION

The mean CQ scores of the four groups are shown in Table 3. The results were subjected

TABLE 2

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS USED IN DERIVATION OF MULTIPLE REGRESSION EQUATION FOR PREDICTING ABSTRACTION SCORE (X') FROM AGE AND VOCABULARY;  $X' = -.28Y + .98Z + 3.82$ .

	Abstraction score (X)	Age (Y)	Vocabulary score (Z)
Abstraction Age		-.41	.64 -.09
M	21.6	34.0	27.8
SD	10.5	13.0	6.5

<sup>3</sup> The author wishes to express his gratitude to Werner Kuhn, Frances King, and Amanda Roth for their help in the collecting and statistical analysis of the data.

<sup>4</sup> Uncorrected for omissions.

TABLE 3  
MEAN SHIPLEY-HARTFORD CQ AND RA SCORES FOR THE GROUPS

Group	N	Conceptual quotient		Residual abstraction score	
		M	SD	M	SD
Acute schizophrenics	30	79.0	14.4	-3.6	7.1
Nonschizophrenic psychiatric	30	81.0	14.7	-2.6	6.6
Chronic schizophrenics	30	71.3	11.7	-6.9	6.9
Normal control	30	86.4	15.7	0.4	7.8

to an analysis of variance.<sup>5</sup> Differences between groups were significant at the .001 level of confidence. Comparing the means of each of the psychiatric groups with that of the control group by means of Dunnett's *t* statistic (Dunnett, 1955) showed the chronic and acute schizophrenics to be significantly inferior to the normal controls at the .01 and .05 levels of confidence, respectively. The difference between the nonschizophrenic and the normal control group is significant at the .20 level. Differences between male and female subjects and the interaction of sex with groups failed to reach statistical significance.

The results thus suggest that schizophrenics obtained lower CQs even though the various groups had been approximately matched for age and vocabulary. The fact that this impairment is much less marked for the acute schizophrenics is consistent with earlier findings (Lewinsohn & Riggs, 1962; Lomont, 1962) showing that acute schizophrenics manifest relatively little impairment on abstraction tasks which are emotionally neutral in content, as the S-H abstraction test is presumed to be.

The mean RA scores of the four groups are also shown in Table 3. An analysis of variance of these scores yielded *F* and *t* values very similar to those obtained for CQ scores.<sup>5</sup> Since the groups were equated for age and vocabulary, this was to be expected. Negative discrepancy scores were obtained by 37% of the normal control subjects, by 70% of the nonschizophrenic patients, by 77% of the acute schizo-

phrenics and by 83% of the chronic schizophrenics. The advantage of the RA score over the CQ is that it is within the age and vocabulary ranges of the population of psychiatric admittees used in this study, independent of the subject's age and vocabulary score. In contrast to this, interpretation of the CQ is confounded by its correlation with these two variables.

The correlation between RA score and Lorr Factor K was -.25 which is significant at the .01 level of significance. Thus the results do not support the second hypothesis and suggest a low but significant relationship between impaired performance on an abstraction test and clinically observed manifestations of conceptual disorganization.

#### CONCLUSIONS

Interpretation of the Shipley-Hartford Conceptual Quotient is confounded by its correlation with age and vocabulary level. A residual abstraction score, which is independent of these two variables, was obtained by computing the difference between the estimated and obtained Shipley-Hartford abstraction subtest score. The mean residual abstraction scores of chronic and acute schizophrenics were significantly lower than those of the normal control group. Also, a low but significant correlation was obtained between residual abstraction scores and interview ratings of conceptual disorganization. The results are interpreted as providing some support for the assumptions underlying the use of the discrepancy between vocabulary and abstraction levels as an index of intellectual impairment.

#### REFERENCES

- <sup>5</sup> A full table of the analyses of variance has been deposited with the American Documentation Institute. Order Document No. 7553 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.
- DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Ass.*, 1955, 50, 1096-1121.
- FECHER, I. B. The efficacy of Shipley-Hartford Re-treat Scale for measuring mental impairment in



- relation to age and other factors. Unpublished master's dissertation, University of Pittsburgh, 1946.
- GARFIELD, S. L., & FEY, W. F. A comparison of the Wechsler-Bellevue and Shipley-Hartford scales as measures of mental impairment. *J. consult. Psychol.*, 1948, 12, 259-264.
- LEWINSOHN, P. M., & RIGGS, ANN. The effect of content upon the thinking of acute and chronic schizophrenics. *J. abnorm. soc. Psychol.*, 1962, 65, 206-207.
- LOMONT, J. F. Schizophrenic thought deficit for problems of anxiety arousing content. Paper read at Midwestern Psychological Association, Chicago, May 1962.
- LORR, M. Multidimensional scale for rating psychiatric patients. *Veterans Admin. tech. Bull.*, 1953, TB 10-507.
- MAGARET, ANN, & SIMPSON, MARY M. A comparison of two measures of deterioration in psychotic patients. *J. consult. Psychol.*, 1948, 12, 265-269.
- SHIPLEY, W. C. *Shipley-Hartford Retreat Scale, Manual of Directions and Scoring Key*. Hartford, Conn.: Hartford Retreat, 1940.
- SHIPLEY, W. C., & BURLINGAME, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.*, 1941, 97, 1313-1325.

(Received November 5, 1962)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 5, 447-449

## OVERINCLUSIVE THINKING IN PROCESS AND REACTIVE SCHIZOPHRENICS

THOMAS S. ELISEO

*Veterans Administration Hospital, Lebanon, Pa.*

Previous research has indicated that schizophrenics show more overinclusive thinking (i.e., failure to exclude behavior unrelated to a task) than do "normals." 17 reactive schizophrenics, 29 process schizophrenics, and 21 chronic general medical-surgical patients, equivalent in vocabulary knowledge and education, were compared on the Epstein Inclusion Test. Contrary to the prediction that the process schizophrenics would commit more errors, no significant differences were found between the groups. It is possible that overinclusion is a characteristic of depression or chronic illness in general, rather than specific to the schizophrenic disorder.

Epstein's study (1953) has supported Cameron's (1944) theory that schizophrenics overinclude more than a normal group. By overinclusion is meant the failure to exclude inappropriate behavior to perform a task properly. Thus, according to Cameron the schizophrenic's conceptual performance is marked by the incorporation of irrelevant ideas and objects. Epstein's operational definition of overinclusion is a paper and pencil test, which presents the subject with a list of 50 stimulus words, each followed by 5 response words and the word "none." The subject's task is to underline all the response words that are absolutely necessary to complete the stimulus words. Epstein predicted and found that schizophrenics underlined more response words than the normals (i.e., overincluded more) and that there were no differences between normals and schizophrenics in the underlining of an insufficient number of response words (underinclusion).

When schizophrenics are classified as either process (Pr) or reactive (Re), according to their

prepsychotic adjustment, differences in their behavior have been noted (Herron, 1962; King, 1958). The Pr, in comparison with the Re, has a less favorable prognosis, shows a deficit in physiological reactivity, gives more "psychotic" Rorschach records, more organic signs and responses of a lower genetic level on the Rorschach, and interprets proverbs on a more concrete level (King, 1958). Additionally, the Pr shows early signs of and a gradual onset of the psychosis and an inadequate prepsychotic social, economic, and sexual adjustment, whereas the Re exhibits a marked precipitating stress, sudden onset of illness, fairly good prognosis, and a comparatively normal prepsychotic adjustment (Fine and Zimet, 1959). Thus, previous research (King, 1958) suggests that it is likely the Pr would show greater deficit than the Re in conceptualization.

The present study is concerned with the overinclusive behavior of Pr and Re. It was hypothesized that Pr would overinclude significantly more than the Re, who would in turn make more

overinclusive errors than general medical-surgical patients suffering from chronic physical illness (GMS) and that there would be no differences in underinclusive errors between the groups.

### METHOD

**Tests.** The Epstein Inclusion Test (Epstein, 1953) and the Word Knowledge Test, which is part of the Airman Classification Battery (Fruchter, 1949), were given to 17 Re, 29 Pr, and 21 GMS. Epstein's procedure was followed in the administration of the Inclusion Test, from which overinclusion and underinclusion scores were obtained. The Word Knowledge Test was given in order to have the subjects equivalent in Vocabulary level. Both tests were administered either individually or in groups of 2 to 4 patients.

**Subjects.** The schizophrenic subjects were selected from a group of newly admitted patients, 18 from each of the following Veterans Administration Hospitals: Tomah, Wisconsin; Battle Creek, Michigan; and Lebanon, Pennsylvania. The 3 hospitals were part of 32 Veterans Administration Hospitals participating in Project VI of the Veterans Administration Cooperative Studies of Chemotherapy in Psychiatry, the purpose of which was to examine the effectiveness of 6 phenothiazines.<sup>1</sup> The patients were newly admitted schizophrenics who had not received any tranquilizing medication for at least 1 month before admission. The majority of the patients were service connected for a schizophrenic disorder. The mean number of hospitalizations for schizophrenic disorders was 3.90 for Pr and 3.29 for Re. Thus, they can be regarded as being chronic, although obviously not "backward" schizophrenics. All subjects were under 55, showing no evidence of organic damage or mental deficiency. A social worker rated the patients on the Elgin Prognostic Scale (Wittman, 1941), as modified by Becker (1959), on the basis of either an interview with the patient or an available social history. Scores on the Becker revision range from 0-50; the median, 25, was selected as the separation point for Pr (scores of 24 or less) and Re (scores of 26 or higher); the subjects with scores of 25 were evenly divided between Pr and Re. Of the original 54 schizophrenics, 18 were classified as Re, 35 as Pr; 1 patient was not classified because of refusal to be interviewed and an inadequate clinical folder. Six of the Pr and one Re either refused or invalidated

the testing. Thus, the final testing involved 17 Re and 29 Pr.

The GMS patients were from Lebanon Veterans Administration Hospital. All suffered from a chronic physical disorder, including diagnoses of heart disease, cirrhosis, pulmonary disorders (e.g., asthma), arthritis, and gastric ulcers (Kessler, 1953). Some of these conditions are considered "psychosomatic"; in the selection of the subjects, those having a concomitant or previous psychotic or neurological diagnosis or showing signs of such were eliminated. The mean for the total number of hospitalizations for the GMS was 3.29. The difference in mean number of hospitalizations for the three groups (Pr = 3.90, Re = 3.29, and GMS = 3.29) was not significant. Comparison of the GMS and schizophrenic groups would provide a closer estimate of the effect of the schizophrenic thought disorder itself, avoiding the confusion of a specific disorder (e.g., schizophrenia) with the effects of a chronic illness in general.

The highest possible score on the Word Knowledge Test was 30. The mean Word Knowledge score was 22.29 for GMS, 20.53 for Re, and 21.41 for Pr, while the mean years of schooling was 10.57 for the GMS, 11.00 for the Re, and 10.69 for Pr. The groups did not differ significantly with respect to vocabulary level and education. The GMS was, however, significantly older ( $p < .01$ ) than the Pr and Re. The mean age for the GMS was 42.95, for Re 35.41, and for Pr 37.71. The product-moment correlation of age and Word Knowledge was .19, indicating little relationship between these variables.

### RESULTS

Because of the unequal sample size, the Kruskal-Wallis  $H$  test was used in the analysis. The correlation between Elgin and Word Knowledge scores was low ( $r = .02$ ). Contrary to the hypothesis, there were no significant differences between the groups in overinclusion (Table 1). The prediction of no differences in underinclusion was supported, although a trend toward greater underinclusion by the GMS is apparent. There was little relationship between Word Knowledge and Overinclusive Scores ( $r = -.07$ ), while an  $r$  of  $-.52$  ( $p < .01$ ) between Word Knowledge and Underinclusive Score suggests that the higher a vocabulary level the less underinclusive errors.

### DISCUSSION

Perhaps the lack of differences between GMS, Re, and Pr can be attributed to the hypothesis that overinclusion is an aspect of depression or of chronic illness in general. It is assumed and gross observation indicates that persons suffering from a chronic physical disorder are usually depressed (Grinker & Robbins, 1954, p. 329). One effect of depression is dif-

<sup>1</sup> The author wishes to thank J. J. Laskey of the Veterans Administration Central Research Laboratory for aid in obtaining schizophrenic subjects and the process-reactive classifications and background data for the subjects. Thanks are due also to D. Pearl, Veterans Administration Hospital, Battle Creek, and V. Pishkin, Veterans Administration Hospital, Tomah, for testing the 18 patients from each of their hospitals. The suggestions and criticisms of L. Aumack, H. Blackburn, and A. Felice of Lebanon were helpful.



TABLE 1  
OVERINCLUSION AND UNDERINCLUSION SCORES  
FOR GMS, Re, AND Pr

	GMS (N=21)	Re (N=17)	Pr (N=29)	Significance test
Overinclusion				
Median	25	27	31	$H = .48$
M	30.71	38.76	41.93	$.80 > p > .70$
SD	21.75	35.90	36.89	
Range	6-70	2-112	1-185	
Underinclusion				
Median	15	10	13	$H = 4.8$
M	18.19	12.00	16.52	$.10 > p > .05$
SD	9.97	7.02	9.69	
Range	5-43	2-29	0-40	

ficuity in concentrating and focusing attention, such that extra effort is required to accomplish a task (Arieti, 1959, p. 425). Epstein (1953) speculated that schizophrenics overinclude because of a defect in attention, resulting in impaired discrimination and generalization. Thus, both depressed and schizophrenic individuals may have difficulty in attending, thereby hampering the ability to perform tasks requiring concentration. A moderately depressed person would be likely, therefore, to make more overinclusive errors, whereas a severely depressed individual would either not participate or if he did would not be responding to the task and would probably underinclude. In the present study no measure of depression at the time of testing was taken. In support of the depression explanation, Payne and Hirst (1957) found that depressives overincluded significantly more than a "normal" nonhospitalized control group.

The finding that the GMS group did not differ from the schizophrenics was surprising and on the surface contradicts, if it is assumed that the GMS is a normal group, most previous research. The usual control groups in studies of schizophrenic thinking have been college students or people not suffering from a major chronic physical affliction. Epstein's control group consisted of nonhospitalized normal individuals from various social and occupational groups—"firemen, recreation groups, housewives, hospital attendants, secretaries, and professional people" (Epstein, 1953, p. 385). Thus, whether schizophrenia or to long-term illness was unknown. The present results suggest that the chronically ill, whether the symptoms are those

of schizophrenia or physical, show some deficit in conceptualization. Whether the deficit is specific to overinclusion or is present in other aspects of thinking between depressives, schizophrenics, the chronically physically ill, and non-hospitalized normals remains to be investigated. In a related, as yet unpublished, study the present writer found that GMS, Re, and Pr did not differ in the interpretation of the figurative and literal meanings of words. Becker (1956) found that newly admitted male Pr and Re differed in the ability to abstract proverbs (the Re made better interpretations than the Pr), whereas the females did not differ. No adequate explanation for this sex difference was found. Thus differences in conceptual deficit between male Pr and Re may be limited only to the interpretation of proverbs.

#### REFERENCES

- ARIETI, S. Manic-depressive psychoses. In S. Arieti (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1959.
- BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 53, 229-236.
- BECKER, W. C. The process-reactive distinction: A key to the problem of schizophrenia. *J. nerv. ment. Dis.*, 1959, 129, 442-449.
- CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944. Pp. 861-921.
- EPSTEIN, S. Overinclusive thinking in a schizophrenic and a control group. *J. consult. Psychol.*, 1953, 17, 384-388.
- FINE, H. J., & ZIMET, C. N. Process-reactive schizophrenia and genetic levels of perception. *J. abnorm. soc. Psychol.*, 1959, 59, 83-86.
- FRUCHTER, B. *The factorial content of the airman classification battery*. San Antonio, Tex.: Lackland Air Force Base, 1949.
- GRINKER, R. R., & ROBBINS, F. P. *Psychosomatic case book*. New York: Blakiston, 1954.
- HERRON, W. G. The process-reactive classification of schizophrenia. *Psychol. Bull.*, 1962, 59, 329-343.
- KESSLER, H. H. *Rehabilitation of the physically handicapped*. (Rev. ed.) New York: Columbia Univ. Press, 1953.
- KING, G. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, 56, 160-164.
- PAYNE, R. W., & HIRST, H. L. Overinclusive thinking in a depressive and a control group. *J. consult. Psychol.*, 1957, 21, 186-188.
- WITTMAN, P. Scale for measuring prognosis in schizophrenic patients. *Elgin state hospital papers*, 1941, 4, 20-33.

(Received November 8, 1962)



TABLE 4

## NUMBER OF CHILDREN IN CLINIC FAMILIES

Number of children	Source (in percentage)			
	Southern clinic	Southern census <sup>a</sup>	Northern clinic <sup>b</sup>	Northern census <sup>c</sup>
1	13	42	15	35
2	32	31	41	33
3 and more	55	27	44	32
Total	100	100	100	100

<sup>a</sup> Census figures refer to metropolitan Jacksonville.

<sup>b</sup> Unknown (3%) were distributed according to observed frequencies of known cases.

<sup>c</sup> Percentages refer to New York State.

As readily seen in Table 5 at the Northern clinic there was a marked incidence of children whose predominant symptom was aggressive behavior. That finding was said to be consistent with the cultural agents contributing to the caseload of that clinic. In the Southern clinic, however, the predominant symptoms were those of nonaggressive behavior such as withdrawal, daydreaming, silliness, and mental retardation. No inferences are drawn on the possibility of cultural agents contributing to these symptom pictures.

## DISCUSSION

The results indicate that the Southern clinic and the Northern clinic serve somewhat different populations. The Northern clinic serves mainly white, middle to high socioeconomic groups; the Southern clinic disproportionately serves whites of middle to low socioeconomic status. While in the Southern clinic, Protestants were underrepresented and Catholics overrepresented, the clientele is more representative of the three different religions in the community than was that of the northern clinic. However,

TABLE 5

## CLASSIFICATION OF DOMINANT SYMPTOMATIC BEHAVIOR AT INTAKE

Dominant symptom	Source (in percentage)	
	Southern clinic	Northern clinic
Aggressive	18	35
Nonaggressive	28	28
Somatic complaints	10	9
Mentally retarded	26	14
Mixed	17	10
Unknown	1	4
Total	100	100

both the Southern and Northern clinics failed to represent the Negro population adequately.

## REFERENCES

- ANDERSON, F. N., & DEAN, H. C. Some aspects of child guidance clinic intake policy and practice: A study of 500 cases at the Los Angeles Child Guidance Clinic, Los Angeles, California. *Publ. Hlth. Monogr., Wash.*, 1956, 16(42), 1-16.
- GINOTT, H. G. Group screening of parents in a child guidance setting. *Int. J. group Psychother.*, 1956, 6, 405-409.
- GINOTT, H. G. Preintake selection of motivated applicants in a community clinic. *Soc. Wk.*, 1959, 4, 105-109.
- HOLLINGSHEAD, A., & REDLICH, F. Social stratification and psychiatric disorders. *Amer. soc. Rev.*, 1953, 18, 163-169.
- MICHIGAN DEPARTMENT OF MENTAL HEALTH. Differential utilization of the facilities of a Michigan child guidance clinic. *Mich. Dept. Hlth. res. Rep.*, 1955, No. 17.
- ROACH, J. L., GURRLIN, O., & HUNT, R. G. Some social-psychological characteristics of a child guidance clinic caseload. *J. consult. Psychol.*, 1958, 22, 183-186.

(Received October 22, 1962)

## METHOD FOR THE STUDY OF STAFF INTERACTION THROUGH RATINGS OF PATIENTS

JEROME FISHER

Langley Porter Neuropsychiatric Institute<sup>1</sup>

Ratings of degree of improvement of 58 hospitalized patients were obtained from a staff of 20 engaged in a therapeutic community type of program. Analysis of staff intercorrelations of improvement yielded an unexpected configuration in which the head nurse of the ward was found to be the star of a sociogram of the staff raters, indicating her importance and influence in many different interactional and interpersonal spheres of ward, hospital, staff, and patient activities. The results of the method applied in this study corroborate earlier observations by Stanton and Schwartz (1954). It is suggested that continuous, systematic evaluations of staff patterns of interaction be made an integral part of therapeutic programs.

In a recent novel, *One Flew Over the Cuckoo's Nest*, Kesey (1962) writes about some of his experiences as a nursing assistant on the psychiatric wards of two California hospitals. Of particular interest is the character of a head nurse, who is called "Big Nurse" and, through the eyes of a psychotic patient, is described as follows:

... she wields a sure power that extends in all directions on hairlike wires too small for anybody's eye but mine; I see her sit in the center of this web of wires like a watchful robot, tend her network with mechanical insect skill, know every second which wire runs where and just what current to send up to get the results she wants [p. 26].<sup>2</sup>

The coincidence between the powers and influence attributed figuratively to Big Nurse and the implications of the results of this study, which was completed about 4 years ago<sup>3</sup> motivated the writing of this paper.

Initially, the purpose was to show how well a staff of 20 people, working together in a type of therapeutic community program (between 1954 and 1959) on a small psychiatric ward of a general medical and surgical hospital, could agree on rating the degree of improvement of 58 male patients. As the study proceeded, how-

ever, a sociometric method was developed for investigating the interactional structure of the staff. It was from these sociometric findings that something approaching the Big Nurse phenomenon evolved.

### METHOD

The patients were representative of veterans who had been hospitalized for various kinds and degrees of psychiatric disabilities. Their characteristics as to age, education, diagnosis, etc., are not relevant to this report and, therefore, are not presented. On the other hand, how well each staff member knew and could recall the patients was a function of how long the patients had been in the hospital. The range of hospital stay was from 1 month to 2.5 years (the latter applied to only one patient). The mean was 5 months.

Among the 20 staff members the following personnel were represented: 2 clinical psychologists, 4 clinical psychology trainees, 1 head nurse, 5 nursing aides, 1 occupational therapist, 1 psychiatrist, 1 psychiatric social worker, 1 recreational therapist, and 4 staff nurses.

The following rating scale was presented to the staff with instructions to rate the degree of improvement shown by each of the 58 patients: 0, Don't know, don't remember the patient; 1, Worse; 2, No change; 3, Some minor improvement; 4, Improved; 5, Considerably improved.

To minimize possible systematic effects of order, the names of the 58 patients were randomly arranged on each of three lists. In turn, the 20 staff members were also assigned randomly to one of the three lists.

After excluding all ratings of zero, a mean, and sigma of the staff's ratings of improvement for each patient were computed. A minimum of four staff ratings for each patient was considered essential for purposes of studying interrater agreement. Four of the 58 patients did not meet this minimum and were eliminated. A subsequent review of their hospital charts revealed that they had been hospi-

<sup>1</sup>From the Langley Porter Neuropsychiatric Institute, Department of Mental Hygiene, California, and the Department of Psychiatry, University of California School of Medicine, San Francisco, California.

<sup>2</sup>From *One Flew Over the Cuckoo's Nest* by Ken Kesey. Copyright 1962 by Ken Kesey. Reprinted by permission of the Viking Press, Inc.

<sup>3</sup>For her assistance in the tabulation and analysis of the data, the author gratefully acknowledges Helen Hudekoff of the Veterans Administration Hospital, San Francisco.

talized a week or less when they left "against medical advice."

### RESULTS

The ratings of the 20 staff members on 54 patients were intercorrelated and only those that reached the .001 level of significance were acceptable. The more conservative probability level was considered appropriate because of the frequency of small numbers of patients rated by some pairs of raters. Each correlation was tested against its own standard error derived for each pair of raters and based upon the particular  $N$  for that pair.

Figure 1 presents a sociogram which, with several modifications in data analysis and graphic illustration, was suggested by Giedt (1957) in his unpublished study of social psychological characteristics of ward treatment teams. It is based upon the statistically significant intercorrelations which were obtained among 10 of the 20 members; each circle represents one of the 10. The lines between circles indicate those with whom each one "agrees" or shares similar points of view. BV, the head nurse, occupies the center position of the sociogram; note that there are nine lines that join her and other staff members. Nearest to her is HN, the occupational therapist, with seven lines. At increasing distances from BV are the other staff members, with DF, a nursing aide, showing only a single line joining him to BV.

### DISCUSSION

By intercorrelating staff ratings of improvement, i.e., their phenomenological judgments about improvement, regardless of real or objective determinants of improvement of the pa-

tients, it is possible to study, among other things, how well the staff agree or disagree among themselves. In addition, the data permit a graphic and sociometric application. The latter could also apply to other scaling techniques,  $Q$  sorts, etc.

The "star" of the sociogram, Figure 1, is clearly BV, the head nurse. Her position of "influence" is underscored particularly by the finding that, of the other nine staff members, only four are on her nursing staff (ZM and HB—nurses; TB and DF—aides); the rest belong to other disciplines (HN—occupational therapist, LR—psychiatrist, CE—social worker, and BJ and FJ—clinical psychologists). Among the latter five, there is considerable interagreement as represented by the mutual lines joining each of them. Yet, each agrees with BV, suggesting the possibility that BV may be responsible for some of their commonly shared points of view about the patients. Another finding of interest is the sociogram position of the psychiatrist (LR). It would seem that his sphere of professional and administrative influence is by no means so extensive as that of BV even though as the only psychiatrist on the ward, it might have been assumed that in his role as chief of the service, he would have been the star of the sociogram. That he was not may have been because he preferred the nonauthoritarian role of a staff member who, along with the others, shared certain prescribed responsibilities, unique to his or her own discipline, for the operation of the total therapeutic program.

The results of a study conducted by Stanton and Schwartz in 1954 are remarkable in their predictive implications regarding the head or charge nurse with respect to this study. They, too, note what seems to be a general tendency to overlook her importance:

What was greatly underestimated was the vital role of the charge nurse as a coordinator. . . . Skill as an intermediary (italics by Stanton and Schwartz) was strikingly necessary for a charge nurse in a position such as we studied. It was essential because a large part of her role was as liaison officer, through whom the activities pertaining to the ward were channeled. She served as a go-between for the various parts of the hospital and patients, as an intra-ward staff coordinator. Finally she was the central figure who either executed, evaded, or altered the arguments and decisions made [Stanton & Schwartz, 1954, p. 157].

Daily morning meetings of most of the staff took place in a large room adjoining the head nurse's office where she reported the events of the preceding day, night, and early morning's

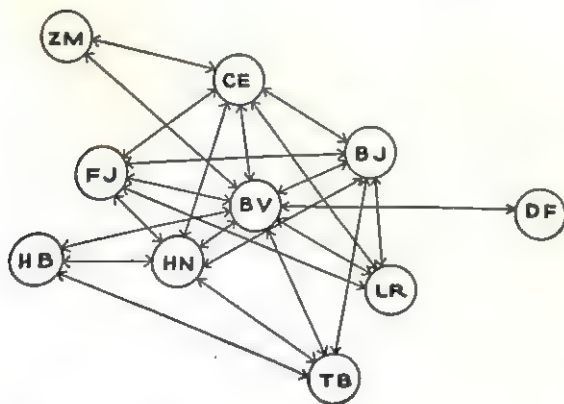


FIG. 1. Sociogram based only upon the intercorrelations of those staff ratings which exceeded the .001 level of confidence.



activities of each patient in the program, as reported to her by the nurses and nursing assistants. In turn she was in the best position to provide "feedback" via impressions and attitudes of the staff to the patients as well as to her night nurses and aides.

Long before this study was undertaken, the staff had begun to examine its patterns of interaction in regular weekly group meetings. As a group, the members of the staff conceived of themselves generally as group oriented, democratic in their decision-making efforts about policy, patient-staff activities, etc. It is interesting, in retrospect, to view the findings against the background of these meetings. Perhaps it is significant that the head nurse and/or one of her staff nurses attended the meetings irregularly because of general ward pressures and patient nursing needs.<sup>4</sup> But when they did attend they were largely silent observers in the group. The aides were also invited, but they rarely, if ever, attended the meetings. One might speculate that had the nursing staff been present and active as participants, these data might have shifted in some interesting and curious ways. Parenthetically, the general impression about these staff meetings was largely supportive of the work of Galioni, Adams, and Tallman (1953), namely,

<sup>4</sup> Besides the psychiatric ward, the nursing staff was also responsible for the nursing care of an adjoining neurology ward of 20 beds. Shortages and high turnover rates of nurses were constant personnel problems. These were some of the reasons offered by the head nurse in explaining the irregular attendance of her and her staff. In light of these factors and circumstances, it is even more curious that BV and some of her staff occupy the position they do in the sociogram.

that treatment of patients is maximized by conducting therapy with the staff, concurrently and continuously. Along these lines, an additional note is suggested by the findings of this study. The phenomenological field represented by the staff is, in effect, the environmental field for the patients. The interpersonal and interactional relations among staff members are assumed to be important variables of therapeutic influence. It would seem necessary, then, that some form of evaluation of roles and activities of the professional staff be built into therapeutic programs so that appropriate assessment measures may be systematically and continuously applied. Results of such studies may prove to "actualize" treatment, as Galioni et al. found, as well to provide more comparative data regarding the study of interpersonal relations among staff members. Further study is planned of related variables, e.g., patient conformity and staff expectations, and their effects on improvement and likeability of patients as rated by staff.

#### REFERENCES

- GALIONI, E. F., ADAMS, F. H., & TALLMAN, F. F. Intensive treatment of backward patients—a controlled pilot study. *Amer. J. Psychiat.*, 1953, 109, 576-583.
- GIEDT, F. H. Patterns of staff attitudes on three psychiatric wards. Paper read at California State Psychological Association, Los Angeles, December 1957.
- KESEY, K. *One flew over the cuckoo's nest*. New York: Viking Press, 1962.
- STANTON, A. H., & SCHWARTZ, M. *The mental hospital*. New York: Basic Books, 1954.

(Received November 5, 1962)

## USE OF THE TERMAN-MERRILL ABBREVIATED SCALE ON THE 1960 STANFORD-BINET FORM L-M ON NEGRO ELEMENTARY SCHOOL CHILDREN OF THE SOUTHEASTERN UNITED STATES

WALLACE A. KENNEDY, VERNON VAN DE RIET, AND JAMES C. WHITE, JR.

*Florida State University*

Consideration was given to the degree of precision which would have been lost had the abbreviated scale only been administered in a normative study of the 1960 Stanford-Binet Form L-M on 1800 Negro elementary school children from five Southeastern states (Kennedy, Van De Riet, & White, 1961). A Pearson product-moment correlation of .99 was obtained between the mental age scores on the full and abbreviated scales for the 1800 Ss stratified according to age, grade, sex, socioeconomic status, and community size, and randomized within these limits. There was little variation from grades 1 to 6, with a mean IQ of 80.7, a SD of 12.4. This low IQ in a homogeneous population is one explanation for the high correlation. With these Ss little precision is gained from the use of the full scale over the abbreviated scale of the 1960 Stanford-Binet Form L-M.

Since the 1937 revision of the test, an abbreviated scale of the Stanford-Binet has been available, using only two thirds of the items and saving one third of the administration time. In large-scale research a saving of about 15 minutes on each subject for individually administered tests is of considerable significance. Although Terman and Merrill (1937), in describing the abbreviated scale, did not present any statistical treatment, they mentioned that an error of up to 10 points might be made in estimating the full-scale IQ from the abbreviated one.

Watson (1951), in his chapter on the Binet, reviews the work from 1937 to 1950, and this source is used by Terman and Merrill (1960) in the manual accompanying the 1960 L-M Revision. In his review Watson (1961) says:

Despite the high correlations found rather consistently, the Terman and Merrill short form . . . has slightly lower mean IQs than those found for the full scale. However, in no case is the mean difference obtained between the short and the full scale statistically significant [p. 261].

Five of the six studies reviewed by Watson (Kvaraceus, 1940; Shotwell & McCulloch, 1944; Spache, 1944; Spaulding, 1945; Wright, 1942) represent highly atypical children mostly mental defectives, or, as in the case of Spache's 100 subjects, nursery and private school children with a mean IQ of 124. Brown (1942), on the other hand, studied 300 randomly selected kindergarten children drawn from 1 year's entering class of a large city. He reported a corre-

lation between the two scales of .94, with a 13.2% of the population differing in IQ by more than 5 points.

From a review of the literature it would seem that there is no large-scale test of the correlation between the two Stanford-Binet scales. It is then little wonder that generally speaking writers of texts on measurement tend to regard the abbreviated scale with great caution, and to speak of it as an emergency expedient only.

In the fall of 1960, shortly after the publication of the 1960 revision of the Binet, the combined Form L-M, the writers conducted a broad, normative study of Negro elementary school children in five Southeastern states. A sample of 1800 subjects, stratified according to age, grade, sex, socioeconomic status, and community size, and randomized within these limits, was tested by a team of examiners from the Human Development Clinic of Florida State University. In addition to the Binet, the California Achievement Test and the Goodenough Draw-a-Man Test were administered.

The results of the normative study, published elsewhere (Kennedy, Van De Riet, & White, 1961), indicate a slightly skewed distribution, with an IQ mean of 80.7, and a SD of 12.4. There were no sex differences in IQ; little difference in IQs of children from urban, metropolitan, and rural areas; but a highly significant relationship between IQ and socioeconomic status; negative correlation with age; and no correlation between grade and IQ.

In looking back over the project some consideration was given to the degree of precision

TABLE 1

MEANS, STANDARD DEVIATIONS, CORRELATIONS, AND STANDARD ERROR OF THE ESTIMATE FOR THE FULL SCALE AND ABBREVIATED SCALE MA OF THE STANFORD-BINET FORM L-M

Chronological age		N	Full scale		Abbreviated scale		Correlation	Standard error of the estimate
Year	Months		MA	SD	MA	SD		
5	60-71	19	61.4	3.8	60.2	4.1	.93	.35
6	72-83	227	67.0	9.4	65.6	9.2	.98	.17
7	84-95	243	74.8	10.8	73.2	10.7	.98	.14
8	96-107	302	84.4	12.1	82.4	12.6	.97	.17
9	108-119	281	93.3	14.4	91.5	15.5	.98	.23
10	120-131	299	103.0	17.3	101.7	18.3	.98	.20
11	132-143	279	110.0	18.5	108.6	19.2	.99	.10
12	144-155	109	109.0	18.4	107.6	18.5	.98	.35
13	156-167	30	98.5	18.5	96.5	19.5	.99	.52
14	168-179	9	106.2	13.9	105.3	12.4	.98	1.00

which would have been lost had the abbreviated scale only been administered. With this in mind, a program was written to compute the revised MA on the basis of the abbreviated scale. It was felt that because of the common CA used in computing the IQ, a comparison of IQs might yield spuriously high correlation and that a comparison of MAs would give a more accurate picture of the power of the abbreviated scale.

Table 1 presents the means, standard deviations, correlations, and standard error of the estimate for the two scales. It can be readily seen that the abbreviated MA mean is 2 points lower than the full form, and that the standard deviation is 6 months lower. The correlations range from a low of .93 for the 19 subjects at 60-71 months, which is equivalent to 5 years, to .99 for the 279 subjects at year 11. The overall correlation was .99.

A correlation of .99 on a sample of 1800 subjects is indeed surprising and it is impressive to note that an error of more than 5 months was made less than 3% of the time, and that an error of more than 10 months was found less than 1% of the time, when estimating the full MA from the abbreviated MA.

It would seem then from this study that when testing Negro elementary school children in the Southeast, the advantage obtained in administering the full scale is slight and the amount of variability between the two IQs is negligible.

## REFERENCES

- BROWN, F. A. Comparison of the abbreviated and the complete Stanford-Binet scales. *J. consult. Psychol.*, 1942, 6, 240-242.
- KENNEDY, W. A., VAN DE RIET, V., & WHITE, J. C., JR. *The standardization of the 1960 revision of the Stanford-Binet Intelligence Scale on Negro elementary school children in the Southeastern United States.* (C.P.R. #954) Tallahassee, Florida: Office of Education, United States Department of Public Health, Education, and Welfare, 1961.
- KVARACEUS, W. C. Pupil performances on the abbreviated and complete New Stanford-Binet scales: Form L. *J. educ. Psychol.*, 1940, 31, 627-630.
- SHOTWELL, A. M., & McCULLOCH, T. Accuracy of abbreviated forms of the Revised Stanford-Binet Scale with institutionalized epileptics. *Amer. J. ment. Defic.*, 1944, 49, 162-164.
- SPACHE, G. The abbreviated Stanford-Binet Scale in a superior population. *J. educ. Psychol.*, 1944, 35, 314-318.
- SPAULDING, P. J. Comparison of 500 complete and abbreviated revised Stanford scales administered to mental defectives. *Amer. J. ment. Defic.*, 1945, 50, 81-88.
- TERMAN, L. M., & MERRILL, M. A. *Measuring intelligence.* New York: Houghton Mifflin, 1937.
- TERMAN, L. M., & MERRILL, M. A. *Stanford-Binet Intelligence Scale.* New York: Houghton Mifflin, 1960.
- WATSON, R. I. *The clinical method in psychology.* New York: Harper, 1951.
- WRIGHT, C. A. A modified procedure for the abbreviated Revised Stanford-Binet Scale in determining the intelligence of mental defectives. *Amer. J. ment. Defic.*, 1942, 47, 178-184.

(Received September 17, 1962)



## BRIEF REPORT

### SOCIAL DESIRABILITY: A REPLY TO WIGGINS<sup>1</sup>

JERALD N. WALKER<sup>2</sup>

University of Washington

A critique by Walker (1962) of an earlier study by Wiggins (1959) was subsequently challenged (Wiggins, 1963) on four main issues: "(a) the actual development of the SD scale, (b) the success of other desirability scales in both studies, (c) the similarity of other control scales in both studies, and (d) Walker's failure to control for the test-retest 'improvement' effect." This paper has demonstrated that these points are inappropriate or inadequate and that Walker's results and conclusions must stand.

Wiggins (1959) reported a study in which he attempted to evaluate the adequacy of a variety of purported measures of social desirability. Walker (1962) questioned his results, and, in a reply, Wiggins (1963) held that the critique was unsuccessful and that his original results must stand. This paper is a reply to Wiggins' criticisms.

The items of the SD scale were originally selected on the basis of perfect consensus among ten judges, rather than on the basis of social desirability scale values, because Edwards (1957) was interested "only in being able to select a set for which there was clear-cut agreement as to what the socially desirable response should be [p. 30]."

There is no inconsistency between the fact that the SD scale was constructed on the basis of a role playing group, and Walker's criticisms. It must be obvious that in a role playing experiment, of two measures of social desirability, the one whose control mean is higher relative to its maximum possible score is bound to come off second best with a scale whose control mean is lower relative to its maximum possible score. (The control mean of Edwards' SD scale was 30.99, and has a maximum possible score of 39; the control mean of Wiggins' SD scale was 12.06, and has a maximum possible score of 40 [Wiggins, 1959].)

<sup>1</sup> An extended report of this study may be obtained without charge from Jerald N. Walker, P.O. Box 787, Presidio of Monterey, California, or for a fee from the American Documentation Institute. Order Document No. 7630 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>2</sup> Now at Human Resources Research Office, Presidio of Monterey, California.

Walker (1962) made no reference to a possible effect of Wiggins' procedure on his control mean for the SD scale. Hence, the similarity of control scores in both studies is not at issue.

With respect to the "improvement effect," Wiggins (1963) reported Windle's (1955) probability levels, rather than the actual mean changes. The latter are very small and cannot account for the greater shifts in Walker's (1962) study. The high test-retest correlations would yield large *t*'s with a very small improvement effect.

Wiggins' (1959) role playing instructions, as originally reported, were clearly a potential source of error. That error existed was well demonstrated by the discrepancy in shifts on the K scale between Wiggins' and other role playing studies, including, of course, Walker's (1962). The question was, "Why didn't Wiggins obtain similar results with the K scale?" It would appear that his instructions were in error.

#### REFERENCES

- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden Press, 1957.
- WALKER, J. N. An examination of the role of the experimentally determined response set in evaluating Edwards' social desirability scale. *J. consult. Psychol.*, 1962, 26, 162-166.
- WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.*, 1959, 23, 419-427.
- WIGGINS, J. S. Social desirability under role playing instructions: a reply to Walker. *J. consult. Psychol.*, 27, 107-111.
- WINDLE, C. Further studies of test-retest effect on personality questionnaires. *Educ. psychol. Measmt.*, 1955, 15, 246-253.

(Received March 25, 1963)

## RELIABILITY OF MEASUREMENT UNDER VARIOUS EXPERIMENTAL INSTRUCTIONS<sup>1</sup>

MARGERY H. KRIEGER AND CHARLES B. SCHULTZ

*Veterans Administration Hospital, Seattle, Washington*

Many studies in clinical and social psychology use repeated measurement designs in which the same instrument (*Q* sort, inventory, semantic differential, etc.) is administered under different experimental instructions (to describe self, ideal, other persons, etc.) on two or more occasions. Some interpolated treatment (e.g., psychotherapy) is often the independent variable. Correlations are calculated between instructions and between occasions to test research hypotheses.

However, the magnitude of such correlations depends on the reliability of the measure. Reliabilities, when reported, are usually based on the instrument used for self-description, and the possibility that reliability may vary with experimental instruction has not been investigated.

This study postulates that test-retest reliability of an instrument used to describe a specific person is greater than that of the same instrument used to describe an abstraction. A 19-scale semantic differential (an instrument described elsewhere [Krieger, in press] in which the scales are balanced for social desirability) was used by 53 undergraduate college students in two classes to describe three specific concepts (President Kennedy, The Class Instructor, Myself) and three nonspecific concepts (A Doctor, A Patient in a General Hospital, A Patient in a Mental Hospital). The choice of nonspecific concepts was determined by interest in their use in another study. Test-retest interval was 1 week.

Data were analyzed in terms of score change. On the 7-point semantic differential scale, test-

retest change can range from 0-6 points. Average change scores were calculated for each scale on each concept ( $N = 53$ ), and used as scores to reflect amount of change. For the entire sample of scales ( $N = 19 \times 6 = 114$ ), these change scores followed a unimodal, symmetric distribution. Mean change score for each of the six concepts was less than one score unit, ranging from Self = .62 to Mental Patient = .89. The  $F$  (between mean square/within mean square) between these six means is 2.67 ( $p < .05$ ). Duncan's new multiple range test (Edwards, 1960) indicates that Self, Instructor, and Kennedy versus Mental Patient, and Self versus Doctor are significantly different at the .05 level or beyond. The overall mean change score for combined specific concepts (.67) is significantly lower than that for nonspecific concepts (.81) ( $t = 3.11$ ,  $p < .005$ ). All specific concepts yielded lower mean change scores than did any nonspecific concept.

These results indicate that test-retest reliability of an instrument varies under different experimental instructions. Specific concepts tend to be more stable than nonspecific ones. The magnitude of any correlations between such concepts will be limited by their reliabilities. Thus, differences in correlations between experimental instructions (often interpreted as dissimilarity) or between occasions (often interpreted as change) may be due to unreliability of measurement rather than to the effect of the independent variable. It should be emphasized that these findings may be limited to the particular instrument used. Repetition of the study with other instruments is planned.

### REFERENCES

- EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960.  
KRIEGER, M. H. A control for social desirability in a semantic differential. *Brit. J. soc. clin. Psychol.*, in press.

(Received December 3, 1962)

<sup>1</sup> An extended report of this study may be obtained without charge from Margery H. Krieger, V. A. Hospital, 4435 Beacon Avenue South, Seattle, Washington, or for a fee from the American Documentation Institute. Order Document No. 7556 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.



## THE RELATION OF THE GORDON PERSONAL INVENTORY TO SEVERAL EXTERNAL CRITERIA<sup>1</sup>

WARREN W. WILLINGHAM

*Harcourt, Brace and World, New York*

AND ROSALIE K. AMBLER

*United States Naval School of Aviation Medicine, Pensacola*

The Gordon Personal Inventory is a self-rating personality scale which measures four traits: Cautiousness, Original Thinking, Personal Relations, and Vigor. The purpose of this study is to determine the relationship between these traits as measured by the inventory and peer judgments of the same traits.

Eleven cadet sections in United States Naval School, Pre-Flight, served as subjects ( $N = 208$ ). The Gordon Personal Inventory (GPI) and nomination forms were administered in a regular testing period in the fourteenth week of training. It is likely that testing approximated administrative conditions since these cadets take similar tests which are used administratively.

After completing the GPI, cadets nominated high and low men on each of four traits corresponding to those measured by the GPI. These traits were defined on the nomination blank by five positive and five negative statements taken verbatim from the GPI. Standard scores for the four traits were derived from the nominations by methods described by Willingham (1959). The number of delinquency reports (disciplinary infractions) each cadet incurred in training was also recorded on the supposition that this variable is related to cautiousness.

The intercorrelations among the GPI scales were lower than those reported in the manual

<sup>1</sup> The opinions and conclusions do not necessarily reflect those of the Department of the Navy.

An extended report of this study may be obtained without charge from Warren W. Willingham, Georgia Institute of Technology, Atlanta 13, Georgia, or for a fee from the American Documentation Institute. Order Document No. 7625 from ADI Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

(Gordon, 1956), but showed approximately the same relative magnitudes. The lower correlations may have been due to restriction in range since each mean was above the fiftieth percentile point on the male college norms reported. The within group reliability for the four nomination scales ranged from .88 to .92. The reliability of number of delinquency reports was .33.

The correlations between the GPI scales and the corresponding nomination scales were: Cautiousness, .38; Original Thinking, .35; Personal Relations, .11; and Vigor, .28. (When  $r = .14$ ,  $p = .05$ ; when  $r = .18$ ,  $p = .01$ .) In no case was there a significant correlation between a GPI scale and an unlike nomination scale. Number of delinquency reports was significantly related to Cautiousness and Vigor as measured by nominations ( $-.30$ ,  $-.15$ ) and by the GPI ( $-.19$ ,  $-.17$ ).

Behavioral manifestation of a trait is the criterion for that trait. Peer nominations and self ratings are surrogate criteria or hypothetical predictors of the behavior. That is, peer nominations are only fairly good criteria for the GPI scales. Within this framework, the scales of Cautiousness, Original Thinking, and Vigor perform moderately well. Personal Relations showed no significant relationship and doubt is cast upon the utility of this scale. The significant correlation of the GPI scales of Cautiousness and Vigor with an extremely unreliable behavioral criterion, number of delinquency reports, speaks well for those two scales.

### REFERENCES

- GORDON, L. V. *Gordon Personal Inventory*. Tarrytown, New York: World Book Co., 1956.  
WILLINGHAM, W. W. On deriving standard scores for peer nominations with subgroups of unequal size. *Psychol. Reports*, 1959, 5, 397-403.

(Received December 17, 1962)



## PERSONALITY VARIABLES IN THE DAP<sup>1</sup>

MARJORIE HEBERLEIN AND F. L. MARCUSE

*Washington State University*

Much importance has been given to the sex of the first figure drawn on the Draw-A-Person test (DAP) as an index of sexual identification. It has been shown that while both sexes tend to draw same-sex figures first, females show more variability. It was thought that there would be personality differences between women drawing same-sex figures and women drawing opposite-sex figures. The present study investigated whether differences in the sex of the figure drawn by women on one or two administrations of the DAP would reflect personality differences as measured by the Edwards Personal Preference Schedule (EPPS).

The DAP was given to introductory psychology classes (Administration A) and 80 women who had drawn females and 80 women who had drawn males were randomly selected. Two weeks later the EPPS was given and 1 month later a second administration (B) of the DAP was given. Four groups of women were isolated: (a) women who drew same-sex figures on administrations A and B (FF), (b) women who drew opposite-sex figures on Administrations A and B (MM), (c) women who drew the same-sex figure on Administration A and opposite-sex figure on Administration B (FM), and (d) women who drew the opposite-sex figure on Administration A and same-sex figure on Administration B (MF).

<sup>1</sup> An extended report of this study may be obtained without charge from F. L. Marcuse, Department of Psychology, Washington State University, Pullman, Washington, or for a fee from the American Documentation Institute. Order Document No. 7524 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

On an a priori (face validity) basis it was decided that the scales of aggression, heterosexuality, and abasement of the EPPS were most relevant. *t* tests were used to compare results from these scales and three significant *t*'s were found, all on the aggression scale. Women drawing males on both or either of the two administrations of the DAP had a higher score on the aggression scale, i.e., had a need for, or showed more aggression.

An analysis of variance was computed to determine whether there were differences in the profiles of the four groups on all the scales of the EPPS. A second analysis was calculated using only the two consistent groups (FF and MM). The only significant *F* ratio found was between scales. These results suggest that there is no consistent profile difference between groups. This finding, however, does not preclude the possibility of a difference on one scale. Phi coefficients were computed for test-retest reliability of the first figure drawn. Small but significant coefficients were found.

In summary, the results of this study gave support to the hypothesis that the sex of the figure drawn on two administrations of the DAP may be indicative of certain personality traits, i.e., there was a greater need for aggression, as measured by the EPPS, found in females who drew males. This drawing behavior was elicited in a culture which, according to Lynn (1959), rewards masculinity, i.e., is androcentric.

### REFERENCE

- LYNN, D. B. A note on sex differences in the development of masculine-feminine identifications. *Psychol. Rev.*, 1959, 66, 126-135.

(Received December 28, 1962)

## NEED FOR SOCIAL APPROVAL AS REFLECTED ON THE TAT<sup>1</sup>

RUSSELL P. NORMAN

*Queens College, Charlotte, North Carolina*

This study was designed to explore the relationships between need for social approval as measured by the Marlowe-Crowne (1960) social desirability (SD) scale, and responses to the TAT. Subjects were 38 female general psychology students who represented high ( $N = 21$ ,  $M = 21.29$ ) and low ( $N = 17$ ,  $M = 6.09$ ) groups on the SD scale. It was hypothesized that in comparison with low SD subjects, high SD subjects would overemphasize socially approved needs (Achievement and Nurturance) and underemphasize socially disapproved needs (Sex and Aggression). It was also predicted that high SD subjects would tell shorter stories, and would recall fewer of the pictures after having taken the TAT.

Ten TAT cards were administered by female examiners using standard instructions. Cards were presented in counterbalanced order to control for position effects on the recall task. Following the TAT proper, subjects were given 2 minutes to recall as many of the pictures as possible. Number of pictures recalled and forgotten and number of words in each story were tabulated for all subjects. The four needs were judged on a 4-point rating scale, ranging from a score of zero (no use of the need) to a score of three (maximum use of the need). Mean interrater reliability was .82.

Data analysis revealed that predictions were upheld for the socially disapproved needs (Sex,  $p < .01$ , and Aggression,  $p < .05$ ). Of the socially approved needs, only Nurturance was significant ( $p < .01$ ), as Achievement failed to

differentiate the two groups. On the whole, it was felt that the hypothesis related to need as expressed on the TAT was supported. The two groups did not differ on length of stories, although as predicted, low SD subjects recalled more of the pictures ( $p < .01$ ).

Definition of Sex and Aggression as socially disapproved needs was somewhat arbitrary, in that one's culture might dictate socially approved outlets for these needs. Although low SD subjects were much more willing to deal with heterosexual themes, practically all subjects, regardless of group, handled this need in a highly conventional and "respectable" manner. Aggression was used much less frequently, possibly because there are fewer opportunities for female subjects to conventionalize these needs.

Concerning the recall of pictures, since high SD subjects tended to avoid the violent and dramatic themes in favor of more passive and "acceptable" stories, the recall value for a given picture would appear to be less than if the subject had actively dealt with a more emotional interpretation. Also, it may have been that the presence of socially objectionable stimuli, whether in the pictures themselves or in the fantasies of the subject, sets repressive defenses into motion which make it more difficult to recall the pictures. One might also infer that low SD subjects were more actively involved at all stages of the task.

The results of this study lend support to the Marlowe-Crowne proposition that SD represents a need for social approval. Furthermore, since these motivational properties have been demonstrated, SD should be studied as a personality variable rather than exclusively in the narrower sense, as a response set operating in the objective testing situation.

### REFERENCE

- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, 24, 349-356.

(Received February 21, 1963)

<sup>1</sup> An extended report of this study may be obtained without charge from Russell P. Norman, Department of Psychology, Queens College, Charlotte, North Carolina, or for a fee from the American Documentation Institute. Order document No. 7627 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

## DIMENSIONS OF PERSONALITY ADJUSTMENT IN MENTAL PATIENTS<sup>1</sup>

JOHN G. HURST

AND

GORDON OLSON

University of California, Berkeley

Anoka State Hospital

An infrequently used source of descriptive information about a patient's adjustive behavior in a mental hospital is possessed by the hospital's nurses and aides. The purpose of this study was to explore the feasibility of using this information to study manifest dimensions of patient behavior, as well as the feasibility of constructing useful rating scales to evaluate these dimensions. A pool of 485 descriptive bits of adjustive behavior was selected from essays written about patient behavior by ward personnel. Seven a priori categories were defined to represent what seemed to be the major functional unities underlying the descriptive bits. Nine expert judges (psychologists and psychiatrists) sorted the bits into these categories. The total pool of phrases was then arranged into a check list which was completed for a sample of 239 patients by an equal number of ward personnel. The Wherry-Winer (1953) method for factoring a large number of items was used to extract the manifest dimensions of adjustive behavior from the responses to the 485 phrase check list. A hierarchical factor structure provided the most satisfactory solution. A general factor, a subgeneral factor and seven group factors were extracted. The resultant factors are briefly described below:

1. *General Hospital Adjustment*. All of the phrases had relatively high loadings on this factor, which was interpreted as a confounding of the raters' general impression of the patient and a halo effect.

2. *Emotional Reactions A: Overt, Hostile, and Other Directed*. This factor, the subgeneral factor, seems to represent overt manifestations of aggression and hostility.

<sup>1</sup> An extended report of this study may be obtained without charge from John G. Hurst, University of California, Berkeley or for a fee from the American Documentation Institute. Order Document No. 7554 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.75 for microfilm or \$2.50 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

3. *Emotional Reactions B: Covert and Self Directed*. The behaviors that define this factor seem to reflect an effort to internalize and control emotion.

4. *Emotional Reactions C: Lack of Contact; Hallucinatory, Delusional, and Bizarre Behaviors*. This factor is represented by behaviors that are markedly inappropriate and lack any obvious direction toward either self or other.

5. *Interpersonal Relations*. This factor seems to assess the patient's willingness to, and potentiality for, relating well to others, both patients and personnel.

6. *Motivation for Rehabilitation*. The behaviors with sizable loadings on this factor reflect a patient who not only has a desire to leave the hospital, but who is also seeking positive means to effect his rehabilitation.

7. *General Mental Alertness*. This factor characterizes a patient who is alert, interested, and comprehending of events and people in his environment.

8. *Responsible and Constructive Activity*. The behaviors that define this factor reflect a patient who actively enters into the hospital program, particularly in maintaining a responsible interest in a work activity.

9. *Supervision and Care*. This factor seems to be related to the amount of staff time or attention required by the patient, other than that resulting from an actual physical handicap. The factors extracted in this study seem to serve as an adequate summary of aides' and nurses' observations of manifest patient behavior as it relates to adaptation to the hospital environment. Thus, if rating scales can be developed which reliably indicate a patient's status with respect to these dimensions of behavior, we could to advantage distill and make use of the large heterogeneous pool of information about a patient that these personnel possess.

### REFERENCE

- WHERRY, R. J., & WINER, B. J. A method for factoring large numbers of items. *Psychometrika*, 1953, 18, 161-179.

(Received March 4, 1963)



## A COMPARISON BETWEEN TAYLOR'S AND FREEMAN'S MANIFEST ANXIETY SCALES

IRIS BALSHAN GOLDSTEIN

*Michael Reese Hospital, Chicago*

Both the Taylor Manifest Anxiety Scale (T-MAS) and the Freeman Manifest Anxiety Scale (F-MAS) are utilized in the assessment of anxiety symptoms. In spite of this, there is no evidence that these instruments measure the same thing. Their intercorrelation of .32, reported by Alpert and Haber (1960), with 40 subjects, leads one to believe that the two scales are indeed different.

The intent of the present study was to investigate the relationship between the Taylor and the Freeman scales and also to see how they were related to the Guilford-Zimmerman Temperament Survey (G-Z).

Eighty women from upper division courses in the Department of Education were given the T-MAS and the F-MAS. There was a correlation of .42 between the two inventories, accounting for only 18% of the variance.

G-Z scores were available for 53 of the women and were correlated with each of the anxiety scales. Among the significant correlations with F-MAS were the scales of Personal Relations ( $-.32$ ) and Friendliness ( $-.36$ ). The significant correlations with the T-MAS were  $-.35$  with Sociability,  $-.55$  with Emotional Stability, and  $-.55$  with Objectivity.

In general it appears that, although they are related, the T-MAS and the F-MAS measure

different aspects of anxiety. Important differences between them become apparent when one looks at their correlations with the G-Z scales. The high negative correlations of the T-MAS with Emotional Stability and Objectivity indicate that the highly anxious person would be characterized by mood fluctuation, worry, and hypersensitivity. With a low score on the Sociability scale, such a person would neither enjoy the company of others nor seek social contacts.

A different group of G-Z scales are related to the F-MAS. A high score on the F-MAS would be typical of the individual who had feelings of contempt and dislike for others. His low scores in Friendliness and Personal Relations would also show a tendency for him to be belligerent and hostile. This agrees with the F-MAS instructions, which state that one is to judge the behavior of others in responding to the questions. In accord with the projective hypothesis, an individual with negative feelings towards others would not think very highly of himself.

The two anxiety scales may correlate with different G-Z traits partly because of differences in item selection, validation methods, and test instructions. It may also be that anxiety is a multidimensional phenomenon and that the two scales tap different dimensions of anxiety. Rather than using one of the two tests to assess levels of anxiety, the experimenter may wish to utilize a combined score on both of them.

### REFERENCE

- ALPERT, R., & HABER, R. N. Anxiety in academic achievement situations. *J. abnorm. soc. Psychol.* 1960, 61, 207-215.

(Received March 11, 1963)

<sup>1</sup> An extended report of this study may be obtained without charge from Iris Balshan Goldstein, Michael Reese Hospital, P & PI, Chicago 16, Illinois, or for a fee from the American Documentation Institute. Order Document No. 7628 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

**Members of the American Psychological Association may  
obtain Professional Liability Insurance protecting  
against *Any* claim**

**... f**or any alleged liability in connection with the rendering, or failure to render, professional services. Our company provides legal counsel by a specialist in the liability field and pays all expenses up to the limits stated in your policy.

Plan developed in cooperation with your Association's Committee on Professional Liability Insurance.

MAXIMUM LIMITS		YOUR COST PER YEAR
PER CLAIM	PER YEAR	
\$ 5,000	\$ 15,000	\$15.00
10,000	30,000	20.25
15,000	45,000	23.25
20,000	60,000	24.60
25,000	75,000	25.65
30,000	90,000	26.55
40,000	120,000	27.60
50,000	150,000	28.35
100,000	300,000	30.90

Available  
exclusively from  
**Smith-Hoggatt-Dawson**  
**Agency**  
CHAMPAIGN, ILLINOIS

**THE SMITH-HOGGATT-DAWSON AGENCY, Champaign, Illinois**

Please send me an application for Professional Liability Insurance.

Name.....

Address.....

City.....State.....

JUST PUBLISHED • JUST PUBLISHED • JUST PUBLISHED

*Prepared under the auspices of the Menninger Clinic*

# CONSTRUCTIVE ASPECTS OF ANXIETY

*Edited by SEWARD HILTNER and KARL MENNINGER*

Seven perceptive "work papers," comparing and contrasting the works of Freud, Kierkegaard, Niebuhr, Tillich, and Augustine—in order to give a contemporary understanding of anxiety in theology, psychology, psychiatry, and related fields.

The contributors are: Ishak Ramzy, Seward Hiltner, Fred Berthold, Jr., Albert C. Cutler, Charles A. Curran, and Paul W. Pruyser.

Foreword by Karl Menninger.

**\$3.50**

Order from your bookstore

## ABINGDON PRESS



## AVOIDANCE OF INAPPROPRIATE SEX-TYPING BY YOUNG CHILDREN<sup>1</sup>

WILLARD W. HARTUP,<sup>2</sup> SHIRLEY G. MOORE

*University of Minnesota*

AND GLEN SAGER

*John Tracy Clinic*

A technique was developed for measuring avoidance of inappropriate sex-typing uncontaminated by sex-appropriate preferences. 69 boys and 78 girls, ages 3 through 8, were observed while playing with sex-inappropriate and neutral toys. Latency of orienting to inappropriate toys was longer for older than younger boys, but no age difference was found for girls. Percent of time spent with inappropriate toys was lower for older than younger children of both sexes. Latencies for boys, but not girls, were longer when E was present than when absent. E's presence did not affect percent-inappropriate scores of either boys or girls. These avoidance measures were not correlated with M-F scores on the It Scale for Children.

The present study is based on the assumption that appropriate sex role behavior in children is, in part, a function of acquired avoidance of inappropriate sex role behaviors. Such avoidance may be acquired by the child via both identification with available models and direct discrimination learning. The important point, however, is that masculinity in boys is assumed to involve avoidance of femininity, and femininity in girls is assumed to involve avoidance of masculinity.

Most empirical studies of sex-typing in children yield data in which the effects of "approach to appropriate-sex objects" and "avoidance of inappropriate-sex objects" are confounded. For example, Rabban (1950) presented children with a group of objects half of which were masculine and half of which were feminine. Rabban's index of sex role identification consisted of the proportion of toys chosen by the child which were appropriate for his sex. Brown's (1956) It Scale for Children (ITSC) is a projective measure of sex role preferences which similarly confounds approach and avoidance tendencies. Selections from among pictures of masculine and feminine objects are made

by the subject for a figure in a drawing rather than for himself. Because of such confounding, most masculinity-femininity tests for children are not adequate devices for studying avoidance of opposite-sex behavior.

The present investigation was designed to assess young children's avoidance of inappropriate objects unconfounded with preferences for sex-appropriate objects. Specifically, avoidance of opposite-sex behavior was studied as a function of chronological age and the presence or absence of an adult examiner.

*Hypothesis 1.* In children between 3 and 8 years, the relation between avoidance of inappropriate-sex objects and chronological age will be positive. This hypothesis is consistent with most theories of identification and also with the assumption that sex-typing stems from direct sex role differentiation by parents. For example, Freud (1933) hypothesized that superego development (which involves sex-typing) is contingent upon the resolution of the Oedipus complex at about 5 years of age. If school-aged children are indeed more identified than preschool-aged children, the former should manifest greater avoidance of inappropriate objects than the latter.

The consequences of intentional sex role differentiation by parents should also be more apparent in older than younger children. Presumably, parental demands to avoid behavior typical of the opposite sex are maintained or increased in frequency during the

<sup>1</sup> The authors wish to acknowledge the assistance of Huda Giddens, Akira Kobasigawa, Gay Meyer, and Aletha Stein in conducting this investigation.

<sup>2</sup> This study was conducted while the first and third authors were at the Institute of Child Behavior and Development, State University of Iowa.

early school years. It seems unlikely that such pressures would be reduced at the beginning of middle childhood.

*Hypothesis 2.* In children from 3 through 8 years of age, there will be greater avoidance of inappropriate-sex activities when an adult is present than when an adult is absent. During the early years, few children consistently self-impose the controls on behavior demanded by society. As Sears, Maccoby, and Levin (1957) suggest, "self-control is dependent on the immediate (or near future) presence of someone who can punish or reward [p. 364]." Although most children eventually become independently masculine or feminine, it is assumed here that early childhood is a transitional period in this regard. During this time the presence of an adult authority figure should stimulate increased conformity to sex role standards.

*Hypothesis 3.* In children from 3 through 8 years of age, the relation between avoidance of inappropriate-sex objects and appropriateness of sex-typing as measured by the ITSC (Brown, 1956) will be positive. This hypothesis is based on the assumption, mentioned previously, that a portion of the variability in scores on M-F tests such as the ITSC derives from individual differences in avoidance of inappropriate-sex activities.

## METHOD

### Subjects

The Ss were 69 boys and 78 girls. The boys, who were enrolled in laboratory schools at the University of Minnesota, were divided into four subgroups: a young nursery school group (YNS) of 17 Ss, ranging in age from 3-2 through 4-4; an old nursery school group (ONS) of 18 Ss, ranging from 4-5 to 5-2; a young elementary school group (YES) of 17 Ss, ranging from 5-5 through 6-11; and an old elementary group (OES) of 17 Ss, ranging from 7-0 through 8-2.

The girls, who attended laboratory schools at the State University of Iowa, were also divided into four groups: a YNS group of 17 Ss, ranging from 3-2 through 4-2; an ONS group of 17 Ss, ranging from 4-3 through 5-4; a YES group of 22 Ss, who ranged from 5-6 through 6-11; and an OES group of 22 Ss, ranging from 7-0 through 8-3.

The schools attended by the Ss typically enroll children of higher than average intelligence from upper middle-class professional and business homes. 95% of the Ss were Caucasian.

Within each of the eight age and sex subgroups,

Ss were randomly assigned to one of two experimental conditions: experimenter present (EP) or experimenter absent (EA).

### Equipment

Laboratory rooms, 10 × 12 feet, equipped with one-way observation facilities were used for observing the Ss. When the child was brought into the room he was led to a toy table located directly beneath the observation window. S was positioned so that he faced the window with his back to E.

Male Ss were presented with an arrangement of feminine and "neutral" toys. The feminine toys consisted of a doll with extra clothes, a mirror and play make-up, pearl beads, a purse, and some dress-up clothes consisting of high-heeled shoes and a hat. The neutral toys consisted of a small pegboard with four pegs, a field guide to butterflies with a small number of pictures and a technical text well beyond the reading level of any of the children, four blocks with heavy duty snaps so that they could be combined in a cluster, and an eight-piece wooden puzzle of a duck with one piece missing. The feminine toys occupied the right-hand two thirds of the table and were separated from the neutral toys by a small space.

Female Ss were also offered two groups of toys, masculine toys and neutral toys. The masculine toys consisted of a pair of boxing gloves, a football, a football helmet, a toy bulldozer, a wheeled tank-transport, a catcher's mitt and baseball, pliers, a razor, a pipe, and tobacco pouch. The neutral toys used with the girls were the same as those used with the boys. The masculine toys were arranged on the right-hand two thirds of the table separated by a small space from the neutral toys.

The neutral toys, in contrast to the inappropriate-sex toys, were deliberately selected for their limited potential for play. They were fewer in number, and were generally in disreputable condition. Thus the dichotomy *inappropriate versus neutral* was confounded with a multidimensional dichotomy *attractive versus unattractive*. It was assumed that preference by a child for the unattractive neutral toys over the more numerous and attractive inappropriate-sex ones would represent deliberate avoidance of the latter materials. No toys appropriate for S's sex were available.

The selection of toys for boys was made arbitrarily by the first two authors. The toys for girls, however, were selected on the basis of ratings by five preschool teachers, six graduate students, five fathers and seven mothers of preschool-aged children. The nine masculine toys used with girls were rated "substantially more appropriate" for boys than for girls by 20 of the 23 raters; the four neutral toys used were judged neutral by at least 19 raters.<sup>3</sup>

<sup>3</sup> Only three of the feminine toys used with male Ss (doll, pearl beads, high-heeled shoes) were included in the ratings. All were judged to be feminine by at least 22 of the 23 raters.



The judges were also asked to rate, on a 4-point scale, the "interest value" of each of the toys. The nine masculine toys selected for use with girls were all judged to be of greater interest to children than any of the four neutral toys. The mean interest value of the masculine toys was 1.6 while the mean interest value of the neutral toys was 2.9, a difference which is significant beyond .001. Thus, attractiveness and sex-typing were confounded in the toy groups for girls as intended.

### Procedure

To begin the laboratory session, *E* positioned *S* at the table opposite the space between the two toy groups. *E* (who was female for both boys and girls) then said:

Here are lots of toys for you to play with. You may play with them anywhere in the room. When you play with them at the table, try to keep these on this side (*E* points to the feminine or masculine toys) and these on this side (*E* points to the neutral toys).

In the EP condition *E* then said, "I'll sit over here while you're playing, and I'll let you know when it's time to go." *E* then sat down and waited, being pleasantly responsive to the child's attentions but never initiating or deliberately sustaining contacts.

In the EA condition *E* said,

I'm going to go right next door and fix this broken toy (*E* indicates material in her hand). You just go ahead and play, and I'll be back in a few minutes.

### Scoring

*Ss* were observed by a concealed *O* once every 15 seconds for a period of 13.5 minutes. A time-lapse light signaled *O* every 15 seconds. Each time the light flashed, *O* recorded where *S* was looking, what *S* was touching, and where *S* was standing. *O* did not look at *S* between observations.

Looking was classified as feminine (or masculine), neutral, or extraneous depending upon the object the moment of observation. Touching was similarly classified depending on the object, if any, that *S* was being handled at the time of the observation. Proximity to toys was scored as feminine (masculine), neutral, or extraneous depending on where the child was judged to be standing in relation to chalked girls on the floor. A 36 × 40 inch feminine (or masculine) area was marked off in front of the sex-typed toys. The neutral area was 36 × 30 inches located in front of the non-sex-typed toys.

Two groups of scores were derived from the observations—latency scores and percent-inappropriate scores. Looking, touching, and proximity intervals consisted of the number of time intervals elapsing before *S* was observed to look at, touch, or come close to an inappropriate toy. Per-

cent-inappropriate scores consisted of the number of intervals in which *S* looked at, touched, or was close to an inappropriate toy divided by the number of intervals spent with all toys.

### Observer Reliability

Two judges observed 41 boys (a total of 2214 observations) and 19 girls (a total of 1026 observations). Agreement (use of the same category to classify a given observation) ranged from 94% to 97% for the latency scores and from 82% to 97% for the percent-inappropriate scores.

### ITSC

The ITSC was administered individually to each *S* within 4 weeks of the session with inappropriate toys by an examiner who had no knowledge of *S*'s behavior in the other session. Administration of the ITSC followed the procedure described by Hartup and Zook (1960).

The ITSC scores were skewed toward the masculine end of the scale for both male and female *Ss*. This skewness was very marked for boys; in fact, so little variability existed in the scores of elementary school boys that it was decided to eliminate these *Ss* from all analyses involving the ITSC.

### RESULTS

Intercorrelations among looking, touching, and proximity scores were computed separately for the latency and percent-inappropriate components. The tau coefficients computed between the various pairs of latency scores ranged from .77 to .79 for boys and from .79 to .84 for girls. Product-moment correlations computed between pairs of percent-inappropriate scores ranged from .85 to .91 for boys and from .78 to .86 for girls.

Since the individual latency and percent-inappropriate scores were so highly interrelated, the main analyses of the data were completed using summary measures. The summary latency score consisted of the mean of the three individual latencies; the summary percent-inappropriate score consisted of the mean of the three individual percent scores.

It was assumed that the latency score would be a positive function of avoidance of inappropriate play materials. Further, the percent-inappropriate score was assumed to be an inverse function of such avoidance. Consequently, it was expected that these two summary scores would be negatively related. This expectation was confirmed, but the



TABLE 1

SUMMARY OF EXTENDED KRUSKAL-WALLIS ANALYSIS OF VARIANCE FOR LATENCY SCORES: MALES AND FEMALES SEPARATELY

Source	<i>H</i>	<i>df</i>
<b>Males</b>		
Among subcells	18.16**	7
Age (A)	9.69*	3
<i>E</i> 's presence-absence (B)	7.01***	1
A × B	1.47	3
<b>Females</b>		
Among subcells	14.28*	7
A	5.83	3
B	2.46	1
A × B	5.99	3

\*  $p < .05$ .\*\*  $p < .02$ .\*\*\*  $p < .01$ .

association is only moderate. For boys,  $\tau = -.48$ ; for girls,  $\tau = -.45$ .

Both summary scores were analyzed, separately by sex, using Treatments × Levels statistical designs. The levels in these analyses consisted of the four age groups; the treatments were *E*'s presence or absence. The assumptions necessary to a parametric analysis of variance could be satisfied only by the percent-inappropriate scores for girls. Both criterion measures for boys, as well as the latency scores for girls, were markedly nonnormal in subcell distribution. Since no effective transformation of these data could be found, the analysis of these scores was conducted by means of an extension of the

TABLE 2

MEAN RANKS FOR LATENCY SCORES OF FOUR AGE GROUPS WITH EXPERIMENTER PRESENT OR ABSENT: MALES AND FEMALES SEPARATELY

	Age groups <sup>a</sup>			
	YNS	ONS	YES	OES
<b>Males</b>				
Experimenter present	34.0	37.3	46.8	48.2
Experimenter absent	15.0	25.0	31.4	42.1
<b>Females</b>				
Experimenter present	30.8	35.2	57.6	44.3
Experimenter absent	37.7	26.1	31.6	43.3

<sup>a</sup> YNS = younger nursery school Ss, ONS = older nursery school Ss, YES = younger elementary school Ss, OES = older elementary school Ss.

Kruskal-Wallis nonparametric analysis of variance.<sup>4</sup>

**Latency.** The results of the analysis of the latency scores are presented in Table 1. First, for boys, the main effects of age were significant ( $p < .05$ ). The association between age and latency is positive, i.e., older boys delayed longer than younger boys in approaching the feminine toys (see Table 2). Mann-Whitney *U* tests between subgroups indicated the following: differences between YNS and ONS and between YES and OES were not significant; latencies were shorter in YNS than in either YES or OES,  $p < .05$  and  $p < .001$ , respectively; latencies in ONS were shorter than in OES,  $p < .02$ .

The main effects of *E*'s presence or absence were also significant for boys ( $p < .01$ ).

TABLE 3

SUMMARY OF EXTENDED KRUSKAL-WALLIS ANALYSIS OF VARIANCE FOR PERCENT-INAPPROPRIATE SCORES: MALE SUBJECTS

Source	<i>H</i>	<i>df</i>
Among subcells	15.89*	7
Age (A)	9.67*	3
<i>E</i> 's presence-absence (B)	3.51	1
A × B	2.71	3

\*  $p < .05$ .

Table 2 shows that for all subgroups, latencies were longer when *E* was present than when she was absent.

Next, for girls, the analysis of latency scores revealed that neither the effects of age nor of *E*'s presence were significant (Table 1). As can be seen in Table 2, age differences in girls' latency scores were in the same direction as the boys'. *H* for main effects of age, however, fell short of significance ( $p < .20$ ).

**Percent-inappropriate scores.** The nonparametric analysis of boys' percent-inappropriate scores revealed significant main effects of age ( $p < .05$ , see Table 3). Younger boys showed more frequent inappropriate behavior than older boys. Mann-Whitney *U* tests between subgroups, as summarized in Table 4,

<sup>4</sup> As developed by Ray Hyman and used by Rosenblith (1959).

TABLE 4

MEAN RANKS FOR PERCENT-INAPPROPRIATE SCORES OF FOUR AGE GROUPS WITH EXPERIMENTER PRESENT OR ABSENT: MALE SUBJECTS

	Age group <sup>a</sup>			
	YNS	ONS	YES	OES
Experimenter present	34.0	35.3	26.8	24.7
Experimenter absent	53.6	47.0	34.6	23.9

<sup>a</sup> YNS = younger nursery school Ss, ONS = older nursery school Ss, YES = younger elementary school Ss, OES = older elementary school Ss.

indicated: the differences between the two NS groups and between the two ES groups were not significant; both groups of preschool boys spent significantly more time with the feminine toys than either group of elementary school boys ( $p \leq .05$ ). The effects of  $E$ 's presence or absence on percent-inappropriate scores was not significant for boys.

The percent-inappropriate scores for girls were analyzed using parametric analysis of variance. The main effects of age were significant ( $p < .005$ , see Table 5).  $t$  tests indicated: the two groups of preschool girls did not differ from each other nor did the two elementary school groups; both nursery groups spent significantly more time with the masculine toys than either of the elementary school groups (see Table 6). The effects of  $E$ 's presence or absence was not significant for girls.

*Relation of ITSC Score to Avoidance of Inappropriate Toys.* Tau coefficients were computed between ITSC scores and the two summary measures of avoidance. These computations were completed separately for pre-

TABLE 5

SUMMARY OF ANALYSIS OF VARIANCE FOR PERCENT-INAPPROPRIATE SCORES: FEMALE SUBJECTS

Source	df	MS	F
Age (A)	3	3174.2	5.22**
$E$ 's presence-absence (B)	1	855.6	1.41
(Cells)	(7)		
$A \times B$	3	982.9	1.62
Cells (w)	64	608.3	
Total	71		

\*\*  $p < .005$ .

TABLE 6

MEAN PERCENT-INAPPROPRIATE SCORES FOR FOUR AGE GROUPS WITH EXPERIMENTER PRESENT OR ABSENT: FEMALE SUBJECTS

	Age group <sup>a</sup>			
	YNS	ONS	YES	OES
Experimenter present	76.5	62.0	37.0	33.0
Experimenter absent	62.3	64.1	53.9	50.7

<sup>a</sup> YNS = younger nursery school Ss, ONS = older nursery school Ss, YES = younger elementary school Ss, OES = older elementary school Ss.

school and elementary school Ss for two reasons: the foregoing analyses revealed age differences in the avoidance measures, and other studies (Brown, 1957; Hartup & Zook, 1960) indicate that ITSC scores are also related to age.

For boys, positive correlations were expected between latency and ITSC scores (since high scores on this scale are equivalent to high masculinity). Negative correlations were expected between percent-inappropriate and ITSC scores. Correlations in the opposite direction were anticipated for girls. Table 7 shows that the direction of each coefficient in the table is as expected, but all are essentially of zero-order.

## DISCUSSION

*Age differences.* The data confirm the hypothesis that, with increasing age, young children increasingly avoid inappropriate-sex objects. With no exceptions, the elementary school Ss spent less time with the inappro-

TABLE 7

KENDALL RANK CORRELATIONS BETWEEN ITSC SCORES AND TWO MEASURES OF AVOIDANCE OF INAPPROPRIATE SEX-TYPING

Subjects	Avoidance measure	
	Latency	Percent-inappropriate
Girls		
Nursery school ( $N = 25$ ) <sup>a</sup>	-.21	.23
Elementary school ( $N = 44$ )	-.07	.14
Boys		
Nursery school ( $N = 35$ )	.07	-.07

<sup>a</sup> Nine nursery school girls were not given the ITSC because the examiner was their nursery school teacher.

priate toys than the nursery school Ss. Also without exception, the latency of initial orientation to the inappropriate toys was longer for the ES groups than for the NS groups. These differences were all significant except for the latency scores of girls. Thus, increasing avoidance of inappropriate sex-typed activities appears to be characteristic of both sexes during early childhood, although manifestations of such avoidance may be more general in the behavior of boys.

An alternate interpretation of the age differences obtained in this study is that "approach to neutral toys" increases with age. The authors reject this interpretation on the grounds that the neutral toys used in the study do not ordinarily attract more interest from older than from younger Ss. The eight-piece wooden puzzle, the plastic pegs, and the four blocks are toys ordinarily provided for preschoolers, not children in the primary grades. The only exception to this generalization is the adult book which was the fourth neutral toy. The volume used, however, was beyond the capacities of any of the children.

Also, the observers reported frequent incidents which suggest that persistent play with the neutral toys is interpretable as avoidance of the sex-typed toys. Comments such as "Those are for girls," and "Those are boys' toys" were frequent. Play with the neutral toys, particularly among the older Ss, was commonly nonvigorous, repetitive, and interrupted on many occasions with side-long glances at the sex-typed toys and *E* (if present). A tendency to "approach neutral toys" may have been operative in the behavior of some Ss but, for the reasons given, it seems most plausible to regard the differences obtained as indications of different degrees of "avoidance of inappropriate-sex toys" rather than "approach to neutral toys."

The findings with respect to age suggest that some factor or factors associated with entrance to elementary school operate to increase avoidance of inappropriate-sex activity. The percent-inappropriate scores of both boys and girls indicate: ES subgroups were more avoidant than NS groups, but neither the two ES groups nor the two NS groups differed from each other.

Pressure to avoid inappropriate-sex activi-

ties is probably increased from a variety of sources about the time middle-class children enter elementary school. Parents express increasing concern about the adequacy of their child's sex-typing; teachers may also introduce demands for sex-typed behavior. School routines are different in nursery and elementary schools. For one thing, cloakrooms and toilets are no longer shared by the two sexes. Also, changes in peer group interactions may contribute to the increased avoidance of inappropriate activities found among ES children in this study. Sociometric studies repeatedly show that an outstanding feature of peer groups in middle childhood (as compared with earlier or later) is cleavage by sex. Consequently, future investigation of avoidance tendencies in relation to sex-typing should include a study of children's demands on each other for appropriate sex-typed behavior.

*Differences Associated with Examiner's Presence.* The hypothesis that the presence of an adult authority figure stimulates increased avoidance of inappropriate-sex toys was confirmed for boys only. Both longer latencies and smaller percent-inappropriate scores were found when male Ss were observed with *E* present. This EPEA difference, however, was significant for the latency scores only.

On the surface, the findings suggest a sex difference in sensitivity to the presence of an authority figure. This difference raises a variety of intriguing questions. Is the process of "internalization" so complete or so rudimentary in young girls, as compared to young boys, that the presence of an adult does not stimulate additional conformity to sex role standards?<sup>5</sup> Or, is a permissive female examiner an irrelevant stimulus in this culture for sex role conformity in young girls? Would the sex difference noted in this study be found if the examiners had been male adults? Parallel effects on young boys and young girls may not occur unless *E* and *S* are of opposite sexes. There is, for example, strong indication in the literature pertaining

<sup>5</sup> The relative levels of avoidance of inappropriate toys by males and females cannot be assessed using the data of this study since the two sexes were not observed with identical toys.



to social reinforcement effects with young children that performance varies according to whether *E* and *S* are of the same or opposite sexes (e.g., Gewirtz & Baer, 1958; Stevenson, 1961). In any event, the effect of *E*'s presence on children's sex-typed behavior is worthy of further investigation. The use of male *Es* is particularly needed before broad generalizations can be drawn from the data of this study.

*Relation of Avoidance Measures to ITSC Scores.* Very low correlations were obtained between ITSC scores and the avoidance measures. Such findings may indicate that sex-typing is not a single dimension in the behavior of young children, an hypothesis suggested by Borstelmann (1961). On the other hand, in the case of girls, the present findings are particularly puzzling. The girls' ITSC scores were heavily skewed in the direction of masculinity. In fact, 29 of the 44 *ES* girls obtained extreme masculine scores. It seems very inconsistent that *ES* girls should obtain such masculine scores at the same time they avoid the masculine toys in the laboratory to a greater extent than the *NS* girls.

Recently, questions have been raised concerning the adequacy of the ITSC as a measure of sex role preferences in girls. For example, Hartup and Zook (1960) found significant increases in femininity on this scale when preschool girls were told that the sexless drawing (for whom *S* make choices) was a girl. Kohlberg and Zigler (1962) also

suggest that the procedures used in the ITSC may be inappropriate for girls. Consequently, the low correlations obtained in the present study may, at least for female *Ss*, result from deficiencies in the ITSC.

# REFERENCES

- BORSTELMANN, L. J. Sex of experimenter and sex-typed behavior of young children. *Child Develpm.*, 1961, 32, 519-524.
- BROWN, D. G. Masculinity-femininity development in children. *J. consult. Psychol.*, 1957, 21, 197-202.
- BROWN, D. G. Sex role preference in young children. *Psychol. Monogr.*, 1956, 70(14, Whole No. 421).
- FREUD, S. *New introductory lectures in psychoanalysis*. New York: Norton, 1933.
- GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behavior for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, 56, 49-56.
- HARTUP, W. W., & ZOOK, ELSIE A. Sex role preferences in three- and four-year-old children. *J. consult. Psychol.*, 1960, 24, 420-426.
- KOHLBERG, L., & ZIGLER, E. The impact of cognitive maturity on the development of sex role attitudes in the years four to eight. Unpublished manuscript, Yale University, 1962.
- RABBAN, M. Sex role identification in young children in two diverse social groups. *Genet. Psychol. Monogr.*, 1950, 42, 81-158.
- ROSENBLITH, JUDY F. Learning by imitation in kindergarten children. *Child Develpm.*, 1959, 30, 69-80.
- SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.
- STEVENSON, H. W. Social reinforcement with children as a function of CA, sex of *E*, and sex of *S*. *J. abnorm. soc. Psychol.*, 1961, 63, 147-154.

(Received December 28, 1962)

## RELATIONSHIPS BETWEEN PERFORMANCE ON THREE TESTS FOR ORGANICITY AND SELECTED PATIENT VARIABLES<sup>1</sup>

PATRICIA J. BRILLIANT AND MALCOLM D. GYNTHY

*Washington University*

This study compared the predictive accuracy of 3 commonly used tests for organicity: Bender Gestalt, Benton Visual Retention, and Graham-Kendall Memory-for-Designs. Each of the 147 patients admitted to the hospital during the experimental period was approached for testing. Complete data were obtained from 120 patients later classified as Chronic Brain Syndrome, Acute Brain Syndrome, Psychosis, Personality Disorder, Chronic Alcoholism, and Other. Results showed that test performance is not related to race or sex, but is significantly related to age, IQ, and education. When the effects of these variables were statistically controlled, performance on all measures predicted the clinical diagnosis of organicity at less than the .001 level of significance. The best single measure was the Bender which, scored by the Hutt-Briskin method, correctly identified 82% of all patients.

Psychologists are frequently asked to assist in the diagnosis of organicity. Among instruments commonly used for this purpose are the Bender Visual Motor Gestalt Test (Bender, 1938), the Graham-Kendall Memory-for-Designs Test (Graham & Kendall, 1946; Graham & Kendall, 1960), and the Benton Visual Retention Test (Benton, 1945; Benton, 1955).

Many investigators have attempted to demonstrate the usefulness of these tests. However, since few of these research workers have used more than one of these tests, little is known concerning the tests' differential effectiveness with a given sample.

Factors other than diagnosis which have been shown to influence performance on these tests are IQ (Aylain & Meltzer, 1962; Griffith & Taylor, 1960), education (Pascal & Suttell, 1951), and age and length of hospitalization (L'Abate, Boelling, Hutton, & Mathews, 1962). However, studies which do not control for these variables are common (e.g., Bowland & Deabler, 1956; Garrett, Price, & Deabler, 1957; Goldberg, 1959; Howard & Shoemaker, 1954; Mehlman & Vatovec, 1956; Niebuhr & Cohen, 1956; Robinson, 1953).

Another important issue involves testing the instruments by means of designs which relate to their clinical use. In other words, psychologists are usually asked to differentiate between organicity and some other (frequently unspecified) psychiatric syndrome, not between organics and normals or between organics and nonpsychiatric patients. However, studies which are otherwise well controlled do attempt precisely this sort of atypical discrimination (L'Abate, Friedman, Vogler, and Chused, 1963; Parker, 1954; Wahler, 1956).

Another point concerns the generalizability of findings. If many restrictions are placed upon one's sample, it is often not possible to generalize beyond the sample obtained. As an illustration, the Pascal and Suttell scoring procedure for the Bender (1951) is restricted to subjects aged 15 to 50 with at least one year of high school. This procedure is thus inapplicable to data obtained from many psychiatric patients. Also, if samples do not match base rates at the hospital involved in the study, generalizations even with local significance are difficult to draw.

The purpose of the present investigation was to use three tests for organicity on a heterogeneous psychiatric sample unrestricted as to age, intelligence, and education and to determine the effects of these variables, as well as those of race, sex, and diagnosis.

<sup>1</sup> This paper is based on a Master's thesis done by the senior author, under the direction of the junior author, at Washington University, St. Louis, Missouri, 1962.

## METHOD

*Subjects*

Every patient admitted to Malcolm Bliss Mental Health Center, St. Louis, during the 6 weeks experimental period was approached for testing. 27 (approximately 20%) of the patients admitted during this time were considered untestable because of extreme belligerence, autism, and/or incomprehensibility. The 120 subjects who completed the test battery were composed of 50 females (22 Negro and 28 white) and 70 males (21 Negro and 49 white). The age range of the sample was 17 to 84 years with a mean age of 40.8 years. Education ranged from 0 to 16 years with a mean of 9.1 years. Nearly all these subjects could be described as falling in the lower socioeconomic class with occupational classifications of unskilled or semiskilled workers.

*Test Materials*

The test materials used in the present study consisted of standard sets of each of three tests: Bender Visual Motor Gestalt, Benton Visual Retention (Form C), and Graham-Kendall Memory-for-Designs. The Vocabulary subtest of the Wechsler-Bellevue, Form I, was also used.

*Procedure*

Every subject was given the test battery within 3 days of admission in order to control for length of hospitalization and degree of medication, both of which may affect test performance. Each subject was given the three tests for organicity, in counterbalanced order to control for practice effects, and then given the Vocabulary subtest. From this latter test, a verbal IQ estimate was obtained. This was subsequently used to correct for the

influence of intelligence on the Benton and Graham-Kendall tests.

After each test battery was given, the tests were scored without knowledge of the patient's diagnosis. For the Benton and Graham-Kendall tests, the standardized objective scoring procedures which yield difference scores corrected for age and intelligence were used. Two scores were obtained for the Benton, a "number correct" and a "number of errors" score. To score the Bender, Hutt and Briskin's (1960) system was used. Subjects were classified as "organic" in terms of Bender performance if they made five or more of the errors of rotation, overlapping difficulty, closure difficulty, cohesion, perseveration, retrogression, angulation difficulty, fragmentation, collision, simplification, impotence, and motor incoordination, all of which have been described by Hutt and Briskin. This scoring procedure was selected over other methods because of its relative simplicity (with practice, about 3 minutes per test protocol) and because of the lack of previous published research on the effectiveness of this technique.

After the test battery was scored, the psychiatric staffing diagnosis was obtained. This was used to assign each subject to one of five diagnostic categories: Chronic Brain Syndrome (CBS), Acute Brain Syndrome (ABS), Psychosis (Psy), Personality Disorders (PD), and Chronic Alcoholism (CA). The diagnostic groupings and their respective frequencies are presented in Table 1. Technically, CA should have been included under the PD category, but it was felt that in these patients there might be subclinical organic involvement sufficient to confound the results if they were included in the PD category. Subjects with the diagnosis of Psychoneurosis ( $N=4$ ), Mental Deficiency ( $N=4$ ), and Undiagnosed ( $N=2$ ) were excluded from analyses involving diagnosis because of the small number of subjects in each category.

TABLE 1  
DIAGNOSTIC CLASSIFICATION OF ORGANIC AND NONORGANIC PATIENTS

Organics	<i>N</i>	Nonorganics	<i>N</i>
Acute brain disorders associated with:		Psychotic disorder	
Intoxication	14	Affective reaction	7
		Schizophrenic reaction	26
Chronic brain disorders associated with:		Personality disorders	17
Circulatory disturbance	4	Chronic alcoholism	26
Convulsive disorder	4	Mental deficiency	4
Senile brain disease	3	Psychoneurosis	4
Alcohol intoxication	5	Undiagnosed	2
Disturbance of metabolism	1		
Unknown cause	3		
Total	34		86



TABLE 2  
MEAN TEST SCORES FOR THE DIFFERENT DIAGNOSTIC CATEGORIES

Diagnosis	N	Bender	Benton (correct)	Benton (error)	Graham- Kendall	Estimated verbal IQ
CBS	19	6.0	4.7	11.9	12.3	89.5
ABS	14	5.3	3.4	6.4	3.5	90.4
Psy	33	3.2	2.7	5.2	2.9	92.5
PD	17	2.1	.5	2.0	1.3	90.5
CA	26	3.3	2.2	3.9	1.9	102.6
All groups	110	3.0	2.7	5.6	3.9	92.6

From the Annual Report of Malcolm Bliss, it was found that the admission rates for the year 1960-61 were CBS, 15%; ABS, 15%; Psy, 35%; and PD, 17%. No base rate was given for CA as a group, since these were included under PD. By testing all patients entering the hospital during the experimental period, an attempt was made to match the experimental base rates with the actual base rates of the hospital. The success of this attempt can be judged by the observed base rates of the experimental sample: CBS, 17%; ABS, 12%; Psy, 28%; PD, 14%; and CA, 22%. Analysis of the difference between the actual and the observed base rates (excluding CA) showed that the two did not differ significantly ( $\chi^2 = 2.80$ ;  $df = 3$ ;  $p > .50$ ); therefore, for the purpose of this investigation, it could be said that the experimental base rates approximated those of the annual admission base rates.

### RESULTS

Table 2 presents the mean scores of the five diagnostic groups for each of the four test measures. On each test the two organic groups performed more poorly (i.e., earned higher scores) than the nonorganic groups. This table also reveals that the separate classification of CA from PD was warranted since, on every test, the mean score of the CA

group was higher than that of the PD group. Combining these two groups would have concealed the true differences between the two. Also included in Table 2 is the mean verbal IQ estimated from the Vocabulary subtest scores for each diagnostic group. The mean estimated verbal IQ for the total sample was 92.6.

The first step in analyzing the results was to consider the possible effects of sex and race upon test performance. Relationships between sex and test scores were found to be negligible ( $r_{p\ bts}$  ranged from .02 to .12), but those between race and Bender and Graham-Kendall scores were significant at the .05 level ( $r_{p\ bts} = .19$  and .21, respectively). Further inspection of the data suggested that these significant relationships might be due to the influence of a third factor, such as age or educational differences between the races. When the effects of age were partialled out, the test score-race relationship became nonsignificant ( $r_{1\ 23} < .17$ ). As these analyses had shown that sex and race were unrelated to test scores, these data were combined for all other analyses.

Table 3 presents the correlations between age, education, and IQ and test scores. Examination of these data reveals that most of the relationships between these variables are significant. This was particularly surprising in the case of the Graham-Kendall and the Benton, as scores derived from these tests are corrected for age and IQ.

Since age, education, and IQ affect test scores, adjustment by analysis of covariance was necessary. The three predictor variables problem presents many statistical difficulties, so reduction to two predictor variables was

TABLE 3  
CORRELATION COEFFICIENTS OF AGE, EDUCATION, AND IQ WITH FOUR DIFFERENT TEST MEASURES

	Bender	Benton (correct)	Benton (error)	Graham- Kendall
Age	.40	.26	.40	.33
Education	-.41	-.02	-.19	-.40
IQ	-.33	.12	.07	-.30

Note. -For  $N = 120$ , an  $r$  of .18 is significant at the .05 level; an  $r$  of .23 is significant at the .01 level.

deemed necessary. IQ was the variable eliminated on the grounds that education and IQ were highly correlated ( $r = .67$  for 118  $df$ ;  $p < .001$ ) and that education was more highly correlated in every case with test scores than was IQ. Table 4 shows the results of the analysis of covariance. (It should be mentioned that adjustments were made only for age on the Benton correct scores as IQ and education were not significantly correlated with these scores.) Inspection of this table shows that all four  $F$  ratios were significant at less than the .001 level.

Table 5 indicates the differential discriminative power of the various test measures. For correct diagnosis of all cases, the Bender seems to be the best single measure (i.e., 82% correct). The Benton error score was least adequate, with only 66% of the total number of cases correctly classified. When only the number of nonorganic patients correctly diagnosed was considered, the combination of the two Benton scores classified 98% correctly. When only those cases correctly diagnosed as organic were considered, the Bender was the best measure with a 67% "hit" rate. The most unsatisfactory result was

TABLE 5  
DIAGNOSTIC ACCURACY (PERCENT CORRECTLY DIAGNOSED) OF TESTS FOR ORGANICITY

	Non-organic diag- nosis	Organic diag- nosis	Total
Bender	92	67	82
Benton (correct)	85	62	76
Benton (error)	95	47	66
Benton (combined)	98	62	81
Graham-Kendall	88	63	78

the number of organics correctly diagnosed by the Benton error score; only 47% were correctly classified. It should be emphasized that in every case the percentage of nonorganic patients correctly classified was much larger than the percentage of organic patients correctly classified.

#### DISCUSSION

The results of this study indicate clearly that performance on three different tests for organicity *does* predict clinical diagnosis of organicity at a very high level of significance. Although the present data show better differentiation of organic from other clinical groups than do the data of L'Abate, Boelling, Hutton, and Mathews (1962), they do not achieve the dramatic discrimination of organics from nonorganics reported by Garrett, Price, and Deabler (1957) who had reported an extremely low misclassification rate. These differences in diagnostic accuracy seem to be a reflection of the differentiation attempted. In the study by L'Abate, in which organics were compared with schizophrenics, the experimental task was more difficult than that usually encountered in a hospital setting: namely, diagnosis of organicity in a heterogeneous psychiatric group. Garrett chose a control group of "normals" (i.e., hospital physicians and administrative personnel) to compare with the organics, a problem which is much simpler than the usual task of the clinician.

For proper evaluation of the predictive powers of tests, the organic base rate of a hospital population must also be considered. In the present study, the base rate was 30%;

TABLE 4

ANALYSIS OF COVARIANCE: TEST SCORES, AGE, AND EDUCATION

Test	Source	$df$	$MS$	$F$
Bender	Adjusted means	4	51.85	41.8*
	Within-groups	103	1.24	
	Total	107		
Benton (correct)	Adjusted means	4	34.50	13.3*
	Within-groups	100	2.60	
	Total	104		
Benton (error)	Adjusted means	4	266.90	79.9*
	Within-groups	103	3.34	
	Total	107		
Graham-Kendall	Adjusted means	4	315.60	11.7*
	Within-groups	103	26.90	
	Total	107		

\*  $p < .001$ .

therefore, a nonorganic diagnosis would be correct in 70% of the cases. Using the Bender, however, one would expect to be correct 82% of the time, thus resulting in a 12% increase in "hits" over what would be obtained by using the base rate strategy. The Benton error score, on the other hand, correctly diagnoses only 66% of the patients, demonstrating that use of a "good" diagnostic instrument may lead to a decrease in diagnostic accuracy!

The most vulnerable feature of the present investigation may be the use of the psychiatric staffing diagnosis as the criterion for organicity, although the reliability of this criterion may have been improved by combining psychiatric diagnoses into a few broad classifications. Unfortunately, alternative approaches to the criterion problem (e.g., EEG) have proved to be as unreliable as psychiatric diagnoses.

The present study might also be questioned for not following actual clinical practice in that objective scoring criteria were used for evaluating the tests of organicity instead of the usual "global" judgments. But Goldberg (1959) has shown that the global judgment method of clinicians is not very accurate. The present study has indicated that, with the Hutt and Briskin (1960) system, a high level of diagnostic accuracy can be obtained with the Bender.

As a final consideration, a practical question may be raised, namely: what are the consequences of "overdiagnosing" or "underdiagnosing" organicity? A review of the present results shows that the four test measures diagnosed the following number of cases as organic: Graham-Kendall, 32 (59 if borderline cases are included); Benton correct, 39; Bender, 45; and Benton error, 68. The comparable figure for the psychiatric diagnosis of organicity was 34.

If we assume that the true number of organics lies somewhere between these two extremes (i.e., 32 and 68), the test measures usually overdiagnosed organicity, whereas the psychiatrist underdiagnosed organicity. The consequences of these "errors" are as follows: the psychological measure will result in a higher percentage of valid positives and a lower percentage of false negatives than the

psychiatrist's; however, the psychiatrists will obtain fewer false positives than the psychological tests. In other words, the tests will predict nonorganicity more precisely than the psychiatrists who, in turn, will predict organicity more precisely than the tests. Perhaps the psychiatrist is the "instrument" psychologists have been searching for to predict accurately the positive occurrence. The combination of psychiatrist to predict organicity and psychologist to predict nonorganicity (both given equal weight) may be the most discriminating approach to the problem.

#### REFERENCES

- AYLAHAN, A., & MELTZER, M. L. The Bender Gestalt test and intelligence. *J. consult. Psychol.*, 1962, 26, 483.
- BENDER, LAURETTA. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. Res. Monogr.*, 1938, No. 3.
- BENTON, A. L. A visual retention test for clinical use. *Arch. Neurol. Psychiat.*, 1945, 54, 212-216.
- BENTON, A. L. *The revised visual retention test: Clinical and experimental applications*. New York: Psychological Corporation, 1955.
- BOWLAND, J. A., & DEABLER, H. L. A Bender-Gestalt diagnostic validity study. *J. clin. Psychol.*, 1956, 12, 82-84.
- GARRETT, E. S., PRICE, A. C., & DEABLER, H. L. Diagnostic testing for cortical brain impairment. *AMA Arch. Neurol. Psychiat.*, 1957, 77, 223-225.
- GOLDBERG, L. R. The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender Gestalt test. *J. consult. Psychol.*, 1959, 23, 25-33.
- GRAHAM, FRANCES K., & KENDALL, BARBARA S. Performance of brain-damaged cases on a Memory-for-Designs Test. *J. abnorm. soc. Psychol.*, 1946, 41, 303-314.
- GRAHAM, FRANCES K., & KENDALL, BARBARA S. Memory-for-Designs Test: Revised general manual. *Percept. mot. skills*, 1960, 11, 147-188.
- GRIFFITH, R. M., & TAYLOR, VIVIAN H. Incidence of Bender-Gestalt figure rotations. *J. consult. Psychol.*, 1960, 24, 189-190.
- HOWARD, A. R., & SHOEMAKER, D. J. An evaluation of the Memory-for-Designs Test. *J. consult. Psychol.*, 1954, 18, 266.
- HUTT, M., & BRISKIN, G. *The clinical use of the revised Bender-Gestalt test*. New York: Grune & Stratton, 1960.
- L'ABATE, L., BOELLING, G. M., HUTTON, R. D., & MATHEWS, D. L. The diagnostic usefulness of four potential tests of brain damage. *J. consult. Psychol.*, 1962, 26, 479.
- L'ABATE, L., FRIEDMAN, W. H., VOLGER, R. E., & CHUSED, T. M. The diagnostic usefulness of two



- tests of brain damage. *J. clin. Psychol.*, 1963, 19, 87-91.
- MEHLMAN, B., & VATOVEC, E. A validation study of the Bender-Gestalt. *J. consult. Psychol.*, 1956, 20, 71-74.
- NIEBUHR, H., & COHEN, D. The effect of psychopathology on visual discrimination. *J. abnorm. soc. Psychol.*, 1956, 53, 173-177.
- PARKER, J. W. Tactual-kinesthetic perception as a technique for diagnosing brain damage. *J. consult. Psychol.*, 1954, 18, 415-420.
- PASCAL, G. R., & SUTTELL, B. *The Bender-Gestalt test*. New York: Grune & Stratton, 1951.
- ROBINSON, NANCY M. Bender-Gestalt performance of schizophrenics and paretics. *J. clin. Psychol.*, 1953, 9, 291-293.
- WAHLER, H. J. A comparison of reproduction errors made by brain-damaged and control patients on a Memory-for-Designs Test. *J. abnorm. soc. Psychol.*, 1956, 52, 251-255.

(Received January 28, 1963)

## UNCONSCIOUS SELF-EVALUATION AS A FUNCTION OF AVAILABILITY OF CUES<sup>1</sup>

RALPH ROTHSTEIN

*Institute for Psychosomatic and Psychiatric Research and Training*

*Michael Reese Hospital, Chicago*

AND SEYMOUR EPSTEIN

*University of Massachusetts*

20 college coeds rated how much they liked each of 4 people judging from the sound of their tape recorded voices presented at 6 levels of distortion. Following this, they rated the voices for how similar they were to their own. Unknown to S, 1 of the voices was her own. At a later date, each S rated her voice after it was identified as such. It was found that Ss rated their own unrecognized voice as most likable when it was neither very distorted nor undistorted. Similarly they reacted most favorably to their undistorted voice when the recognized similarity was moderate in degree. Opposite to findings on unrecognized voice, Ss reacted unfavorably to their recognized voice.

In comparison to the rather voluminous research dealing with the phenomenal self, relatively little experimental work has been done on nonconscious aspects of self-judgment. Wylie (1961) points out that those studies dealing with the unconscious or non-phenomenal self have typically been vague and incomplete and have not yielded a clear pattern of results. This she attributes, in part, to methodological problems and the difficulties inherent in establishing construct validity when working with unconscious phenomena.

One approach to the study of unconscious self-attitudes which seems promising surprisingly has been employed by only three investigators, (Epstein, 1955; Huntley, 1940; Wolff, 1933; 1943). Introduced by Wolff in 1933, the technique utilizes judgments by people of others' and their own expressive movements. In all three studies it was found that Ss rated their own unrecognized self-productions differently from how they rated others. Wolff (1943) and Huntley (1940) reported that self-ratings in the absence of recognition tended to be extreme in both directions. They hypothesized that as disguise of an expressive movement increases, self-rating becomes increasingly extreme. The hy-

pothesis follows from the assumption that unconscious determinants play an increasing role as disguise increases. Epstein (1955) found no evidence for a group of self-under-evaluators, and attributes Wolff's and Huntley's findings to a statistical artifact. He found the typical reaction to an unrecognized self-production to be one of over-estimation, and that increasing the disguise decreased the amount of overestimation. The studies of self-attitudes referred to utilized many expressive forms, e.g., voice, handwriting, photographs, etc. presented at a single level of disguise. It would thus not be determined to what extent the judgments were influenced by degree of disguise apart from the form in question. A more desirable approach in studying the relationship between self-judgments and level of recognition is to determine how judgments change as self-related cues are experimentally varied within a single expressive form. This is done in the present study by presenting voice samples at various levels of distortion. In addition to investigating "unconscious" self-judgments as a function of disguise, the present study also makes comparisons with conscious ratings. Voice was selected because it offers important cues for forming judgments of others, lends itself to experimental manipulation, and is not readily

<sup>1</sup> The study was done while Rothstein was at the University of Massachusetts.

recognized when presented as a tape recording of someone else.

## METHOD

### Subjects

Twenty female college students enrolled in the introductory psychology course at the University of Massachusetts served as Ss.

### Procedure

In order not to arouse suspicion about the nature of the study, Ss were informed that the purpose of the investigation was to determine the relationship between quality of voice and personality. The Minnesota Multiphasic Personality Inventory was administered to make the instructions convincing. Samples of voice were obtained by having S read two simple and affectively neutral paragraphs into a microphone so that tape recordings could be made.<sup>2</sup> When recordings were completed S was requested to return in 2 weeks to serve as a judge of the voices of others. Voice recordings were presented at all levels of distortion to be described below, and S rated her own (without knowing it) and the voices of three others for "favorableness" and then for "similarity recognition." Approximately 1 month later, S listened to and evaluated her own voice after being told it was her own.

### Procedure for Obtaining Stimulus Materials

Ratings were obtained for three levels of distortion of voice presented in the natural manner (Voice Forward) and for three levels of distortion in which the tape was played backwards (Voice Backward). To obtain the Voice Forward levels, the following procedures were used:

*Voice Forward (VF)*. The playback of Ss tape recorded voice was made as clear as possible, using a Pentron model 9T3C tape recorder.

*Voice Forward Moderately Distorted (VFD<sub>1</sub>)*. The tape was re-recorded through a microphone onto another tape. As a result of the double recording, fidelity was reduced.

*Voice Forward Considerably Distorted (VFD<sub>2</sub>)*. The same procedure as in VFD<sub>1</sub> was repeated a second time. However, during the second re-recording, volume was maximally increased and range of frequency maximally restricted.

The following procedures were employed to obtain the Voice Backward levels:

*Voice Backward (VB)*. Voice recordings were played in a background direction, making the reproduction unintelligible and sounding like a strange language.

<sup>2</sup> The paragraphs read by the subjects were adapted from Memory for Stories (The Wet Fall), Year 8, Form L (Terman & Merrill, 1937).

*Voice Backward Moderately Distorted (VBD<sub>1</sub>)*. Beginning with VB, the same procedure was followed as for VFD<sub>1</sub>.

*Voice Backward Considerably Distorted (VBD<sub>2</sub>)*. Beginning with VBD<sub>1</sub>, the same procedure was followed as for VFD<sub>2</sub>.

Preliminary to the experiment proper, voice levels were presented to a group of 10 graduate students who rated them on degree of distortion. There was perfect agreement for both Voice Forward and Voice Backward.

### Measurement of Unconscious Self-Judgment

Two weeks after recordings were made, S was presented with four voices to judge, one of which, unknown to her, was her own. Following the procedure used by Huntley (1940), measures were obtained by the use of a graphic rating scale. The scale was anchored at 6 points with labels varying from "dislike extremely" to "like extremely." Instructions were given to use the entire sensitivity of the scale by drawing a line across an appropriate point on the scale, rather than simply checking an internal. All four voices were presented so that S could develop a frame of reference before any ratings were made. The voices were then presented for S to rate immediately following each. A reminder was given to rate the person, and not the voice, per se. Since each S rated and was rated by four other Ss, the data could be arranged in a series of 4 × 4 matrices in which the rows represented raters and the columns voices. In one case, the self-rating was the first presentation, in another the second, in another the third, and in another the fourth, thereby controlling for position effects. VB preceded VF, and order of presentation was always from greatest to least distortion. The S was not informed that actually the same four voices were presented to her for ratings, under different conditions of distortion.

It was decided at the beginning of the experiment that if S spontaneously recognized her voice while rating it for favorableness, her data would be discarded. That no such spontaneous recognition occurred is probably a result of the strong set produced by informing Ss that they would rate the voices of others, as well as a result of the small work sample used.

### Measurement of Similarity-Recognition

To help insure that only unconscious self-ratings were included, as well as to obtain information on the relationship of self-evaluation and level of recognition, ratings for "similarity recognition," as described by Epstein (1955), were obtained. After completing ratings for favorableness, S was asked to rate how similar each voice was to her conception of her own. A rating scale anchored at six points was used for this purpose, the range of which varied from "opposite" to "identical." It had been decided that if a rating of identical oc-



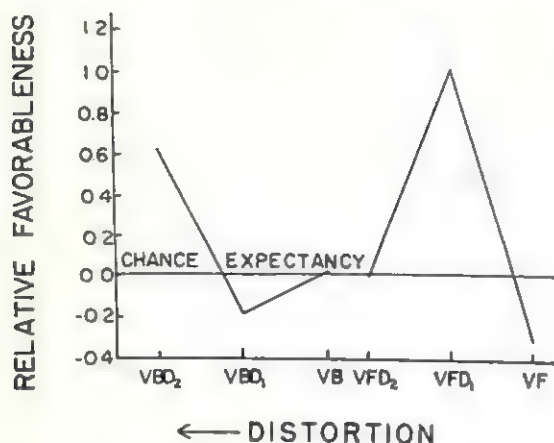


FIG. 1. Relative (self minus other) mean favorableness scores for all levels of distortion.

curred, *S* would be questioned to establish whether she had recognized the voice when rating it for favorableness. Only one rating of identical occurred and this at the  $VFD_1$  level. When questioned, *S* said that she had not recognized it previously, but did when directed to look for similarity to her own.

#### Procedure for Obtaining Conscious Self-Judgments

To determine whether there were any differences in *Ss* conscious and unconscious self-evaluations, approximately 1 month after unconscious self-ratings were made, *S* was told which voice was her own, and asked to rate it when it was presented in an undistorted manner. The purpose of the month delay was to allow for completion of other aspects of the experiment before giving away the fact that one's own voice was included, which might have been communicated to prospective *Ss*. If *S* indicated on questioning that she had heard her recorded voice at some time prior to the experiment, she was instructed to indicate what her reaction had been then as well as what it was now. A check list with five choices varying from "very displeased" to "very pleased" was used for recording judgments. Finally, *S* was asked to write a brief description of her reaction to her voice sample following identification.

### RESULTS

#### Favorableness of Unconscious Self-Ratings

Graphic ratings were converted into scores by using a ruler containing 25 divisions to measure the distance along the scale to *Ss* judgment. The use of so refined a scale insured that limitations in the score would be restricted to the judgments themselves, and not to the selection of too coarse a unit of

measurement. While judgments are not expected to approach the precision of the scale, there is nothing to lose by using an excessively refined scale, so long as *S* is not required to make equally fine distinctions. To obtain scores of unconscious self-judgment, which involves a relative comparison between self and other ratings for each of four voices ( $4 \times 4$  array of ratings), the means of the rows and columns as well as the array were calculated. The formula used for this determination was:  $\text{Self Rating} + M_r - M_c - M_e$ . Subtracting row and column means control for rating bias and quality of voice, respectively.

In Figure 1, it can be seen that the undistorted presentation of Voice Forward (VF) tends to elicit unfavorable reactions to self, that the ratings become favorable with the first level of distortion, and neutral with the second. The only significant departure from a mean of zero, as determined by *t* tests, occurs at the  $VFD_1$  level, where significance is reached at the .05 level. There is considerable random variability in the ratings of Voice Backwards, no level of which approaches significance. Plotting ratings for self and others separately, as in Figure 2, reveals that as distortion increases favorableness decreases, and this holds both for VF and VB. Also the combined ratings for VB are more unfavorable than for VF. Apparently, despite instructions to rate the person and not the

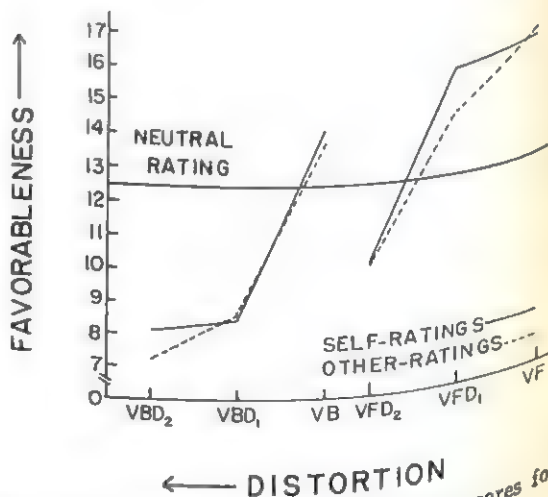


FIG. 2. Mean favorableness ratings in raw scores for self and others for all levels of distortion.

voice, there is a tendency to confound the two, and to judge the unfamiliar less favorably than the familiar. It is of some interest that the ratings of the least distorted level of VB are more favorable than the most distorted level of VF, although the former is more distorted than the latter. Apparently, separate frames of reference are established within each of the forms of presentation.

### *Extremes of Unconscious Self-Rating*

To investigate extremeness of self-ratings apart from direction, the variance of the favorableness scores was compared at the different levels of distortion. The higher the variance, the more the self-ratings differ in both favorableness and unfavorableness from the ratings of others. It is found that with increasing distortion self-ratings became more uniform, i.e., less extreme. Differences between levels ( $VFD_2 - VF$ ;  $VBD_2 - VB$ ) however fall short of significance,  $p = .09$  and  $.07$ , respectively. When the variability of self- and other-ratings were determined separately one finds that at the undistorted level (VF) ratings of self are more extreme both in favorableness and unfavorableness than ratings of others (Self  $\sigma^2 = 36.60$ ; other  $\sigma^2 = 17.89$ ). But again this difference just falls short of significance,  $p = .07$ .

### *Similarity-Recognition*

In Figure 3 it can be seen that for VB presentation, Ss are better able to perceive the similarity of their own and other voices to their actual voice when distortion is decreased and that self-similarity ratings are consistently, though slightly higher than ratings of others. For VF on the other hand while there is a sharp increase in recognition of one's own voice, the perceived similarity of other voices, to one's own, levels off after an initial increase. Analysis by  $t$  test reveals a significant difference between self- and other-ratings for the  $VFD_2$  (.05 level) and VF (.01 level) presentations.

It is noteworthy that despite a general inverse relationship between distortion and similarity recognition,  $VFD_2$  obtains a lower similarity recognition score than VB, which

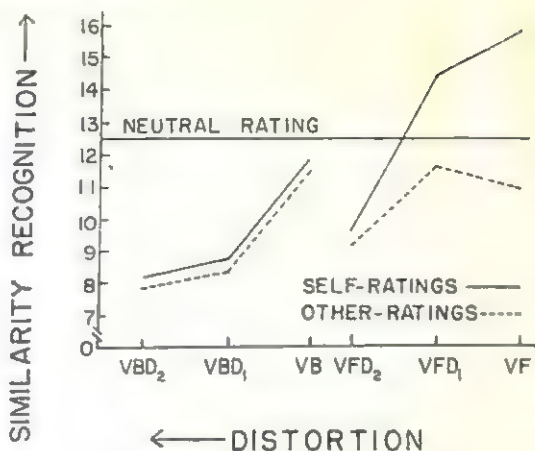


FIG. 3. Mean similarity recognition ratings in raw scores for self and others for all levels of distortion.

suggests, as with favorableness ratings, that a separate frame of reference is established within the two modes of presentation.

In order to determine the relationship between unconscious self-judgment and degree of perceived similarity of voice to one's own, Ss were divided into three levels of similarity recognition for each voice level. Following Epstein (1955), a low similarity recognition group was composed of Ss whose self-ratings indicated "no" or "very little" similarity and who gave another voice a rating of higher similarity than their own, a high similarity recognition group was composed of Ss whose self-ratings indicated "much" similarity and who gave no other voice as high a similarity rating as their own, and a moderate similarity recognition group was composed of Ss not falling in either of the other groups. A simple randomized analysis of variance for favorableness ratings as a function of level of perceived similarity produced significance at the .05 level for the undistorted voice. At this level Ss rated their unrecognized voice most favorably when recognition was only moderate. In order of increasing similarity recognition, the mean scores for favorableness are as follows:  $-2.50$  ( $N = 3$ ),  $.72$  ( $N = 13$ ),  $-1.85$  ( $N = 4$ ). For the  $VFD_1$  level of presentation, no significant relationship is obtained; at all other levels, there were too few cases of high similarity recognition to allow for reliable comparisons.



### Conscious Self-ratings

Eighteen of the original 20 Ss took part in this phase of the experiment. Of the written descriptions of reaction to identified voice, 4 were favorable, 10 unfavorable, and 4 neutral. Thus there is a tendency for Ss to react negatively to their recognized voices while reacting favorably to their voices when unaware that it was their own. While it is not possible to make direct comparisons between mean ratings on the conscious and unconscious self-ratings, as the wording on the scales is different, the data can be dichotomized into favorable and unfavorable ratings on both scales. Of nine Ss who performed differently on the two scales, eight made more favorable judgments of unidentified than identified voice, which is significant at the .05 level. When the very first reaction to hearing one's recorded voice is taken into account, there are eight Ss who differ on judgments of recognized and unrecognized judgments, and in all eight cases judgments of identified voice are less favorable than of unidentified voice, which is significant at the .01 level.

### DISCUSSION

The results fail to support the supposition of Wolff (1943) and Huntley (1940) that with greater disguise of a self-production, defenses are increasingly circumvented, and judgments of self become more extreme. Findings, while falling short of statistical significance, suggest the opposite trend, as increasing disguise results in less extreme reactions to self, both of a favorable and unfavorable nature. It was found however that self-judgments were significantly more favorable than judgments by and of others at the level of moderate disguise. Furthermore, for the undisguised presentation of voice, self-judgments differed from other-judgments in being more positive only when recognition of similarity to one's own voice was moderate. The curvilinear relationship between disguise and personality revealingness is not surprising, for while disguise may reduce defensiveness, it provides fewer cues which can stimulate reactions at any level,

i.e., with sufficient disguise, the stimulus becomes meaningless.

In this respect, the findings in the present study are consistent with studies of projective tests (Kenny, 1954; Kenny & Bijou, 1953; Murstein, 1958) which have reported that more revealing material is elicited by TAT cards of moderate than of low or high ambiguity. The completely negative findings on VB suggest that this form provides too few self-related cues to be a useful stimulus. Significant results using this form in another study (Epstein, 1955) may relate to differences in the population of the two studies.

The findings on conscious self-judgment are consistent with the frequent observation that people are disappointed when hearing their recorded voice. Yet the question posed is why one should react favorably to one's unrecognized voice but unfavorably to one's recognized voice. One plausible explanation is that Ss are less guarded or inhibited about expressing narcissistic, i.e., favorable feelings about self in response to self-related cues, when they are unaware that the production is actually their own. But at the undistorted level, recognition of similarity to one's own voice increases to a point where it is possible that Ss may have come to suspect self-inclusion. At this level, ratings become unfavorable. And when conscious recognition did occur, reactions to self were strikingly unfavorable, Ss revealing much displeasure with their voices. The negative self-reactions associated with conscious recognition is interpreted as reflecting culturally acquired modesty about appearing vain and narcissistic. In future work it would be desirable to investigate in greater detail correlates of both favorable and unfavorable reactions to self in order to clarify the meaning of these response tendencies.

### REFERENCES

- EPSTEIN, S. Unconscious self-evaluation in a schizophrenic and control group. *J. abnorm. soc. Psychol.*, 1955, 50, 65-70.
- HUNTLEY, C. W. Judgment of self based upon records of expressive behavior. *J. abnorm. soc. Psychol.*, 1940, 35, 398-427.



- KENNY, D. T. Transcendence indices, extent of personality factors in fantasy responses and the ambiguity of TAT cards. *J. consult. Psychol.*, 1954, 18, 345-348.
- KENNY, D. T., & BIJOU, S. W. Ambiguity of pictures and extent of personality factors in fantasy responses. *J. consult. Psychol.*, 1953, 17, 283-288.
- MURSTEIN, B. I. The relationship of stimulus ambiguity on the TAT to the productivity of themes. *J. consult. Psychol.*, 1958, 22, 348.
- TERMAN, L. M., & MERRILL, M. A. *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence*. New York: Houghton Mifflin, 1937.
- WOLFF, W. The experimental study of forms of expression. *Charact. Pers.*, 1933, 2, 168-176.
- WOLFF, W. *The expression of personality*. New York: Harper, 1943.
- WYLLIE, R. C. *The self-concept*. Lincoln: Univer. Nebraska Press, 1961.

(Received February 11, 1963)

## DOGMATISM AND SENSORY ALIENATION: AN EMPIRICAL INVESTIGATION

MARTIN F. KAPLAN

*University of Iowa*<sup>1</sup>

AND ERWIN SINGER

*City College of the City University of New York, School of Education*

This study was designed to investigate certain aspects of the concept of "alienation" as employed by Fromm. Using Fromm's framework and suggestions offered by Schachtel concerning inherent autocentrism and allocentrism in various sensory modalities, it was hypothesized that: (a) dogmatism as measured by Rokeach's scale and adequacy on sensory discrimination tasks would vary negatively with each other and (b) differences between dogmatic and nondogmatic individuals would be more pronounced in their performances<sup>4</sup> utilizing predominantly autocentric modes than on tasks involving essentially allocentric modalities. Sensory discrimination was assessed by requiring Ss to match and/or differentiate between stimuli in 5 areas. 13 dogmatic and 13 nondogmatic Ss (controlling other relevant variables) served as populations. Data obtained supported the hypotheses at least at the 5% level.

### THEORETICAL BACKGROUND

Since the days of Breuer and Freud (1957) the notion that psychopathology is characterized by the patient's strenuous attempts to remove from consciousness certain stimuli, be they internal or external, has been prominent in most theories which claim some affiliation to what is loosely referred to as "depth psychology." From "repression" and the other "mechanisms of defense" through "dissociation" to the concept of "estrangement" the common denominator is the assumption that humans block from awareness what they deem noxious. What has changed in the evolution of psychological theory are the constructs concerning what is blocked, how the blocking takes place, and why humans engage in such a process of reduction of consciousness. Through the years the emphasis has shifted, at least in certain quarters, from postulating man's attempt to defend himself against libidinal incestuous and/or aggressive inner stimuli to the hypothesis that there exists a more or less prominent tendency to remove from awareness the fact of one's

very being and the sense of responsibility implicit in the awareness of being or existing. The empirical research to be reported here addresses itself to some aspects of this debate by investigating certain operations presumably underlying the latter position, i.e., the position which suggests that the central problem to be studied is the degree to which some people attempt to minimize the very awareness of themselves and to what extent this minimization is intricately associated with certain character trends.

As is well known, in discussing psychopathology Fromm (1941, 1955) takes his point of departure from the concept of alienation as developed by Marx in his *Economic and Philosophic Manuscripts* (Fromm, 1961). Alienated or estranged man is to Fromm essentially Kafka's Mr. K. or Camus' Stranger. The process of estrangement is said to proceed via a diminution or outright elimination of awareness of the self as an active and separate organism, leading to a sense of being which is characterized by passive drivenness and lack of responsibility for action, be that intellectual, emotional, or even physical activity. Alienated man is said to lose his sense of being "subject" and to become more and more "object" in his own eyes. Schachtel (1959), without actually employing the term alienation, deals with the

<sup>1</sup>This report is based in part upon a Master's thesis by the senior author at the Graduate School of Education of the City College of the City University of New York. The work was aided by a grant from the Max Gewirtz Fund. This support is gratefully acknowledged.

concept indirectly in his discussion of "objectification" and what he calls the "two basic perceptual modes," the allocentric and the autocentric modes. He suggests that the various sensory modalities are more or less capable of functioning in one or both modes. He proposes that the olfactory and gustatory senses operate essentially in the autocentric modes, i.e., are relatively incapable of objectifying the world but rather arouse reactions and feeling tones which bring about some position vis-a-vis the stimulus yet contribute "no clue at all about object structure [p. 88]." On the other hand, sensory modalities such as hearing and vision are much more capable of objectifying. Inner reactivity may but does not necessarily occur within the perceiver when these latter sensory modalities are employed. A person may correctly identify a given visual stimulus as green without assuming a personal position in terms of a pleasure-displeasure continuum to the stimulation. Schachtel points out that this differential functioning of sensory modalities "is not a matter of absolute alternatives, but rather of the relative predominance of objectification [p. 98]."

If the utilization of the predominantly autocentric modes enhances man's immediate self-awareness through attention to personal experience in addition to mere attempts in objectification; and conversely, if rather exclusive focusing upon predominantly allocentric modalities minimizes subjective experience through emphasis upon grasping the nature of the object without attending to subjective reactions, then it would follow that self-alienated man would de-emphasize the former, i.e., autocentric modalities, and in doing so reduce subjective experience and awareness of self. The predominantly allocentric modes, however, should not undergo such a de-emphasis in the self-alienated person.

While employing different terms Fromm's (1961) discussion of idolatry is consistent with this formulation. He states at one point:

This is, incidentally, also the psychology of the fanatic. He is empty, dead, depressed, but in order to compensate for the state of depression and inner deadness, he chooses an idol, be it the state, a party, an idea, the church, or God. He makes this idol into the absolute, and submits to it in an absolute way [pp. 44-45].

If we pursue the implications of Fromm's thinking further it would become reasonable to hypothesize that the fanatic in his "empty, dead, depressed" state would emphasize the allocentric modalities and de-emphasize the autocentric modes or to restate the proposition in more empirically verifiable terms, that increasing fanaticism or dogmatism would be associated with decreasing autocentric perception and experience. If the fanatic is self-estranged then his self-estrangement is likely to proceed primarily via alienation from those sensory modalities which bring in their wake self-awareness or perhaps require some basic readiness for self-awareness, the awareness of self as a separate, living, and immediately reacting unit.

Before discussing the specific experimental procedures employed a theoretical point needs clarification. While fanaticism and dogmatism are not necessarily synonymous, they seem closely related, for the fanatic as well as the dogmatic are essentially individuals who are in the jargon of psychological description "rigid," incapable of tolerating ambiguity and in Rokeach's (1960) terms "closed-minded."<sup>2</sup> But closed-mindedness in its conceptual meaning is not restricted to the "fanatic" or the "dogmatic" but permeates in one form or another all manifestations of psychopathology. For example, when one says that a person is relatively devoid of insight one proposes, in effect, that the individual in question is closed to or unaware of his feelings, tendencies, impulses, or reactions, is in the broadest sense of the word self-alienated so we return to our original theoretical point of departure, i.e., the proposition that pathology is essentially characterized by a type of fanaticism or dogmatism which reflects itself in varying types of closed-

<sup>2</sup> Even though Rokeach makes a subtle distinction between rigidity and dogmatism, employing the former in relation to specific behavioral manifestations, the latter in connection with total belief systems or general orientations, this distinction is for our purposes unnecessary. We shall employ the terms dogmatism and rigidity interchangeably, both denoting a schematized and fixed outlook and orientation towards self and the world. Consequently, the term rigidity as employed here comes close to what Rokeach refers to as dogmatism or closed-mindedness.



mindfulness and rigidity, these rigidities expressing themselves in different forms, but all of these forms detracting from self-awareness.

### HYPOTHESES

In order to investigate this proposition empirically we hypothesized that the degree to which the tendency towards dogmatism is operant within a given individual to that extent self-awareness should appear reduced, this reduced self-awareness operationally reflecting itself in decreased emphasis upon the autocentric sensory modalities and relatively elevated emphasis upon the allocentric modes of perception. Schachtel's discussion suggests a rather specific hierarchy or continuum along the autocentric-allocentric dimension. In ascending order of allocentric potential he arranges the sensory modalities in approximately the following order: olfactory, gustatory, tactile, auditory, and visual. Thus he suggests that the olfactory modality is hardly capable of bringing about objectification and is closely associated with the autocentric mode of perception, the visual modality is most capable of objectifying and therefore potentially most closely associated with the allocentric mode of perception (1941, 1955).

On the basis of these considerations the following specific hypotheses were developed: 1a. Highly dogmatic individuals should exhibit significantly lowered sensory acuity when compared with individuals who are relatively free of dogmatism; 1b. As a logical corollary of the above hypothesis there should exist a negative correlation between dogmatism and sensory acuity; 2. Finally, this differentiation in sensory acuity should be most prominent in the autocentric sensory modalities and less pronounced in the allocentric modes. Put a bit differently, declining sensory acuity is expected in all human beings on the allocentric-autocentric continuum but this decline will be more pronounced in highly dogmatic than in relatively nondogmatic individuals.

### METHOD

For purposes of this investigation dogmatism was operationally defined as the score achieved by a

subject on the Rokeach Dogmatism Scale (1960, pp. 71-100). Previous researches have established the usefulness of this device by finding satisfactory correlations between scores achieved on the Scale and personality variables serving as criteria measures. By and large it may be considered the best existing objective measure of closed-mindedness and a host of studies support the reliability and construct as well as criterion validity of the instrument (1960, pp. 101-108, 171-274).

The assessment of the sensory acuity or sensitivity to stimuli proved a more complicated matter. Two procedures were employed to assess *olfactory sensitivity*. First five odors from Henning's Odor Prism were selected. These odors were along the dimension of spiciness at known distances from each other. Three spices, one at a time, were placed under the nose of each subject while his eyes were closed and following MacDonald's suggestion (Woodworth, 1958) the subject was asked to compare the second stimulus presented with the first and the third and to judge which of these resembled most closely the second stimulus. Each set of three spice stimuli was presented in systematically rotated order so that each spice became at least once a comparison stimulus. In the second *olfactory task* three toilet waters were employed. Each scent was presented, one at a time and then one of the three was re-presented, the subject being asked to indicate with which of the original stimuli he was now confronted. Systematic rotation of order of presentation was observed and care was taken to make each scent a comparison stimulus.

Three sets of salt, sugar, and vinegar solutions, each set containing a heavy, a medium, and a slight concentration of the respective additives, served as the stimuli for the *gustatory task*. The concentration in the "medium" solution, however, was always closer to one of the extremes rather than exactly in the middle between them. Each set of the three solutions was presented for three trials, presentation varying from trial to trial. The subject was asked to sip each solution without swallowing and after each sip he was to rinse his mouth with warm water. Each trial consisted of the presentation of two of the three stimuli and the subject was then given the third solution and was asked to indicate which of the preceding two the third solution resembled more closely.

Three sets of swatches of cloth, each 6 inches square, each set containing three swatches of similar texture, were employed in testing for *tactile acuity*. The texture of one set was rough, of the other smooth, and of the third silky. Even though the nature of the materials in each set was rather similar they were of different weave. Subjects were asked to avert their eyes and were given one swatch at a time to touch. They were to rub one hand over the material but were not allowed to lift or to pinch it. 10 seconds were allowed for each swatch. A trial consisted of the presentation of all three materials constituting a set and the re-presentation of one of them, the subject being asked

to judge which of the three he had received for the re-presentation. Each set was presented employing systematic rotation procedures three times, making for nine trials in all. Care was taken that the test stimulus appeared at varying points in the presentation procedure.

For the *auditory task* a record, "The Measure of Your Phonograph's Performance," manufactured by the Dubbings Company was secured. The record has several bands, each with a different tone of known frequency. These bands were grouped into three sets, each set containing three tones. Again, the frequencies of two tones were closer to each other than to the third tone. Each set was presented three times, varying the order of presentation, and after the subject had listened to two tones he was asked which of the two the third stimulus resembled more closely.

Finally, for the *visual task* sample paint cards (2 × 1.5 inches) were obtained. There were 12 cards, each a different tint of green, pink, or yellow. In each trial two tints of the same color were presented and the subject was then asked to judge (without being shown the original stimuli again) to which of the two the third tint seemed closer. Since the paint company had ranked the tints in order of distance from the primary color, there was an objective basis for judging a response correct or incorrect.

In order to make scores achieved on the various tasks comparable the number of trials were held constant from task to task, allowing one point for each correct identification or judgement. In this fashion the possible range of scores remained the same for each sensory modality.

From a population pool of 40 randomly selected subjects who had been given the Rokeach Dogmatism Scale, those 13 who had scored highest and those 13 who had scored lowest on the Scale were given all the above-described sensory discrimination tasks. Procedures for both comparison populations, personnel carrying out the testing and similar relevant variables were controlled and kept constant.

## RESULTS

It will be recalled that we suggested in Hypothesis 1a that highly dogmatic individuals would exhibit significantly lowered sensory acuity when compared with individuals who are relatively free of dogmatism. Data having bearing upon this assertion are presented in Table 1.

The data speak for themselves. On five of the six measures evaluated the low dogmatic group proved superior to the highly dogmatic individuals to a degree which makes the differences observable rather unattributable to chance. Only on the visual task is the difference observed, although mov-

TABLE 1  
MEANS AND STANDARD DEVIATIONS OF ALL MEASURES  
EMPLOYED FOR THE HIGH DOGMATIC GROUP AND  
THE LOW DOGMATIC GROUP

Measure		Low dog- matic N=13	High dog- matic N=13	Differ- ence	t <sup>a</sup>
Dogmatism scale	M	127.92	174.62	-46.72	10.06***
	SD	5.99	13.23		
Olfactory task	M	5.31	3.70	1.61	3.18***
	SD	.98	1.47		
Gustatory task	M	6.00	4.46	1.54	2.79**
	SD	1.36	1.35		
Tactile task	M	5.46	4.00	1.46	2.70**
	SD	1.51	1.11		
Auditory task	M	6.61	5.46	1.15	1.93*
	SD	1.30	1.60		
Visual task	M	6.77	6.07	.70	1.24
	SD	1.25	1.49		
Total sensory	M	30.15	23.69	6.46	5.41***
	SD	3.23	2.55		

<sup>a</sup> *p* values are for one-tailed tests.

\* *p* < .05.

\*\* *p* < .01.

\*\*\* *p* < .005.

ing in the predicted direction, relatively small and hence likely to have been the result of chance factors and artifacts. The data must therefore be considered as highly suggestive that Hypothesis 1a represents a valid assertion.

In Hypothesis 1b, a logical corollary of Hypothesis 1a, the above findings are simply expressed in different statistical fashion. A Pearson product-moment correlation of  $-.61$  was calculated between our measures of sensory acuity and scores achieved on the Rokeach Scale. This value is attributable to chance factors less than 1:1,000 and, as already stated, expresses in different fashion the observation that with increasing dogmatism sensory acuity decreases.

It was also predicted that the differences in sensory acuity between the two groups would not be uniform from modality to modality but much more follow a gradient of differentiation from relatively minor differences in the primarily allocentric modes to marked differences in the primarily autocentric modes (Hypothesis 2).

Once again the results seem rather un-



equivocal in support of the hypothesis. The gradient of differences is apparent even though neither group exhibited the exact pattern of performance predicted. It will be noticed that both groups did better on the gustatory task than on the tactile task but this is irrelevant for it should be recalled that absolute performance on the tasks is less important than the gradient of differences in performance predicted and in this respect the data are evidently conclusive. The particular order of differences observed is attributable to chance only on a 1:120 basis, or maximal for the present design.<sup>3</sup>

### DISCUSSION

At this point we wish to mention certain methodological difficulties. It is well nigh impossible to construct tasks in varying sensory modalities which may be deemed equal in difficulty. We have no way of telling whether the visual tasks were comparable to the gustatory tasks. Of course, it would be possible to systematically eliminate items here and add items there until a series of tasks is evolved on which a criterion population would exhibit equal competence from task area to task area and having established such a scale, administer it to various pre-selected populations. But we believe that this would be begging the question, for we still would not know whether the tasks were comparable; we would only know that a given population obtained identical scores on all tasks. We therefore decided not to concentrate so much on the development of items of illusory comparability in difficulty but much more upon the investigation of differences between groups on various types of tasks and the question of whether these differences would vary in accordance with certain theoretical propositions. Our focus of inquiry was in this direction and we conclude that the closed and dogmatic person appears estranged or alienated from his sensory impressions and this estrangement is more significantly observable in the predominantly autocentric modalities than in those modalities which

lend themselves more readily to what Schachtel has called "objectification."

Our data suggest two additional observations and comments. First of all they tend to make one doubt the frequently made assertion that there exists an inherent contradiction between the so-called "objective" and "subjective" approaches to experience. On the contrary, our low dogmatic individuals did not only perform in superior fashion on the predominantly allocentric tasks but also on the predominantly autocentric test items, and vice versa, the highly dogmatic, while doing almost as well as the low dogmatic on the allocentric items, did a remarkably inferior job on the autocentric procedures. In other words closed-mindedness and open-mindedness are not phenomena which appear in some areas of an individual's living and not in others, but much more they are pervasive phenomena, the close-minded is not only closed-minded vis-a-vis personal and highly autocentric experience, he is also close-minded vis-a-vis objectifiable stimulation. Conversely, the open-minded person exhibits this orientation vis-a-vis both autocentric experience and objectifiable data. As a matter of fact the idea suggests itself pointedly that the very openness to experience which makes it possible for a person to be *au courant* with his sensations also enables him to adequately objectify such sensory impressions. And this openness to sense impressions apparently runs parallel to openness to ideas, willingness to examine them critically, and careful analysis of thought. This should not be too surprising for the history of man and civilization is replete with incidents and examples in which those who have contributed most significantly in one area have also shown themselves open to the pursuit of other seemingly opposite endeavors. Leading scientists have been remarkably receptive to aesthetic experience, eager to follow fantastic intuitive hunches, and have exhibited genuine interest in humanistic intellectual pursuits; humanistic intellectuals and men prominent in creative artistic endeavors have exhibited astounding interest and insight in matters scientific. The "two cultures" which Snow (1959) is fond of talking about may not be as remote from each other as one might suspect for genuine

<sup>3</sup> We are indebted to Jacob Cohen of New York University for suggesting the appropriate statistical treatment of this datum.



involvement in either may very well require a personal quality essential for both.

Furthermore, if neurosis is viewed as dogmatism of sorts, i.e., a restriction of horizons and rigidification along certain lines, these particular lines depending upon the theoretical bent of an author, then our results are highly supportive of the hypothesis that estrangement or alienation is markedly associated with such rigidification, be this association causal, concurrent, or resultant. The design of the present study throws little light upon this question of sequence. We can only speculate and suggest that the observation that the autocentric modalities which appear in human development more prominently at early stages do not lend themselves to the type of objectification our culture demands in order to "get along." It is readily conceivable that such cultural demands bring in their wake a more or less dogmatic orientation towards existing, an orientation which, as we

have shown, also entails sensory alienation. Further research bearing upon this sequential issue is sorely needed.

## REFERENCES

- BREUER, J., & FREUD, S. *Studies on hysteria*. New York: Basic Books, 1957.
- FROMM, E. *Escape from freedom*. New York: Farrar & Rinehart, 1941.
- FROMM, E. *The sane society*. New York: Rinehart, 1955.
- FROMM, E. *Marx's concept of man*. New York: Frederick Ungar, 1961.
- ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.
- SCHACHTEL, E. The dynamic perception of the symbolism of form. *Psychiatry*, 4, 1941, 79-96.
- SCHACHTEL, E. *Metamorphosis*. New York: Basic Books, 1959.
- SNOW, C. P. *The two cultures and the scientific revolution*. New York: Cambridge University Press, 1959.
- WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1958.

(Received February 15, 1963)

## THE DIMENSIONALITY OF THE MMPI CLINICAL SCALES AMONG NORMAL SUBJECTS

PAUL DEMPSEY

*University of California, Davis*

Scales 1-9 and K were tested for dimensionality in separate samples of 40 males and 40 females. The test used was based on contextual analysis, a method that yields an explicit factor score for each S, thus permitting a practical coefficient of dimensionality:  $r_d$ , the Pearson  $r$  between scale scores and contextual scores derived from the same items. Results were highly similar in both samples, only Scale 5 (Mf) showing a difference in dimensionality between sexes. Differences between scales were marked. Scales 1 (Hs) and 7 (Pt) showed excellent dimensionality; 3 others, 8 (Sc), 9 (Ma), and K, were fairly satisfactory; the rest showed coefficients too low to warrant treatment as interval scales among normals. The best scales can profitably be shortened; the better scales are improvable. The poorer appear suitable mainly for nondimensional uses in configural work and profile interpretation.

One striking gap in MMPI research is the absence of any explicit investigation of the dimensionality of scale scores on the clinical scales. There have been several factorial studies of the intercorrelations of scale scores, many of which have yielded similar results (Welsh, 1956), but these have perforce assumed the dimensionality of the scales they were investigating. There has also been a set of factorial studies of the intercorrelations of the items on single scales (Comrey, e.g., 1958a, 1958b, 1958c), but while these have not made the *a priori* assumption of the dimensionality of the scales, neither have they provided any basis for testing it. In short, apart from the descriptions provided by McKinley and Hathaway on the original construction of the scales (Welsh & Dahlstrom, 1956), there is available no direct information on this important question.

The need for filling this gap in research arises primarily because current use of the MMPI requires of the clinical scales that they make consistent differentiations among normal subjects and among hospital patients as well as that they distinguish normal subjects from hospital patients, and both of these from specific criterion groups. In a word, current use of the MMPI calls for scales that make dimensional distinctions *within groups*, although the clinical scales themselves were constructed by methods that insure such distinctions only *between groups*. As Hathaway (1956) describes the general process of

scale construction, all items were selected on the basis of group comparisons, in his words, "by contrast of criterion groups with other clinical groups and with various normal groups [p. 104]." The process of cross-validation also rested on group comparisons, in this case on the extent to which the frequency distributions of scale scores in samples of the criterion abnormal type were separated from the frequency distributions shown by other kinds of samples. In brief, although various groups were defined, no attempt was made during scale construction to get within-group differentiation that reflected systematic individual differences.

The success of the configural approach (Meehl & Dahlstrom, 1960) and the uniformity of factorial results (Welsh, 1956) provide positive if indirect evidence that at least some of the scales are making consistent differentiations within at least some groups of subjects at least some of the time. Nonetheless, there can be considerable advantage in making explicit the degree of dimensionality of the various clinical scales *within* the main groups or subpopulations in which the MMPI is commonly used to make distinctions among persons. As a first approximation, this can be taken to mean within normal samples and within samples of hospital patients with miscellaneous diagnoses. The present paper is concerned with the dimensionality of the scales among normal subjects.

To test the dimensionality of any scale re-

quires both a definition and a method. An evaluation of the various indexes of dimensionality outlined by Lumsden (1961) indicates that historically there has been relatively little basic disagreement in definition but considerable variation in method. Unfortunately none of the methods has been really satisfactory; whether because of limitations in logic, the need for excessive calculations, or difficulty of interpretation, no one of the suggested indexes has recommended itself for general use. The index of dimensionality used in the present study was devised by the writer as an improvement over existing indexes. It was made possible by the availability of contextual analysis, a recently developed method of isolating the major dimension or factor underlying the responses of a given sample of subjects to a given set of stimuli (Dempsey, 1962; Dempsey & Baumhoff, 1963). For present purposes, the important distinguishing characteristic of contextual analysis is that it yields specific numerical positions or factor scores for all subjects, scores that exist only in theory in the centroid analysis of items. Since these scores, called contextual scores, order the subjects with respect to the major systematic individual differences in their performances and are unaffected by influences not related to those differences, they make possible a simple and convenient definition of a perfect unidimensional scale. *A perfect unidimensional scale is one in which there is a perfect correlation between scale scores and contextual scores derived from the same items.* This definition is logically compatible with the general definitions suggested by Loevinger (1947), Lazarsfeld (1950), Lord (1952), and Lumsden (1961), as well as with Guttman's definition of a quasi-scale (Guttman, 1950). It also states one extreme value of a very simple and readily interpretable index of dimensionality. That index, called the major coefficient of dimensionality,  $r_a$ , is simply the Pearson product-moment correlation coefficient between scale scores and contextual scores. The major coefficient, whose possible range is from .00 to +1.00, is to be interpreted like any Pearson  $r$ . Its availability makes testing the dimensionality of the original MMPI scales eminently practicable.

## METHOD

### Subjects

Two samples of 40 Ss each were used. All members of both samples were classifiable as normal, although there was considerable variation in adequacy of adjustment. The Ss were selected so that there was homogeneity within samples with respect to sex, age, and educational level, but important differences between them in all three variables. One sample was composed of male graduate students with a median age of 27 years.<sup>1</sup> The other was composed of female undergraduates with a median age of 19 years. The samples are called the male and female samples on the assumption that sex is the most important distinguishing variable.

### Procedure

The MMPI protocols obtained from the Ss were scored on Scales 1-9, and also on K. The entire analysis was carried on *within* samples. A total of 20 separate tests of dimensionality were made, one in each sample for each of the above scales. The index used was the major coefficient of dimensionality  $r_a$ , described earlier in this paper. The steps followed in making the tests were identical for all scales and both samples. Brief descriptions of the chief steps are given below.

1. *Preparation of Agreement Matrix.* Paired comparisons were made of the response patterns of the Ss to all items of a given scale, each possible pairing receiving an agreement score equal to the number of identical responses. These scores were then assembled in a matrix in which both rows and columns designated persons.

2. *Contextual Analysis of Agreement Matrix.* This technique has been described in some detail elsewhere (Dempsey & Baumhoff, 1963). It is basically an iterative summing process that quickly converges on a unique solution for any matrix in which there are systematic differences among agreement scores. The solution provides for each S a dimensional position or factor score that is maximally consistent with the major systematic differences appearing in the agreement matrix. These scores are called contextual scores.

3. *Calculation of the Coefficient of Dimensionality.* The Pearson product-moment method was used to correlate the scale scores of the Ss on a given scale with the contextual scores derived from the same items. The scale scores were raw scores, and no K correction was added. Tests of significance were made where appropriate.

<sup>1</sup> The writer wishes to express his appreciation to Donald W. MacKinnon, Director of the Institute of Personality Assessment and Research, University of California, Berkeley, for permission to use the protocols of these subjects.



RESULTS

The 20 coefficients of dimensionality calculated for the 10 MMPI scales in the two samples of normal males and normal females are shown numerically in Table 1. The range of the  $r_d$  coefficients for the male sample is from a high of .96 for Scale 7 (*Pt*) to a low of .11 on Scale 3 (*Hy*). For the female sample the range was approximately the same, the highest  $r_d$  coefficient being .91, also on Scale 7 (*Pt*), and the lowest being .01 on Scale 5, the *Mf* Scale. On 7 of the 10 scales the coefficients for the male and female samples are within .08 of one another, on 2 the difference is .22, and on 1 scale, Scale 5 (*Mf*), the male sample shows a coefficient of .61 and the female sample .01. This last difference is significant ( $p < .01$ ). Since the male sample shows a slightly higher coefficient on each of the other scales as well, there is a small but significant sex difference in the overall height of the coefficients of dimensionality.

For two of the scales, Scale 7 (*Pt*) and Scale 1 (*His*), the  $r_d$  coefficients in both samples exceed .90. On these, scale scores account for more than 80% of the variance attributable to systematic (dimensional) differences in the *Ss'* performances. Three other scales show coefficients high enough to account for 50% of the dimensional variance: Scale 8 (*Sc*), where  $r_d = .84$  and  $r_d = .79$  in the male and female samples, respectively; Scale 9

(*Ma*), where  $r_d = .82$  and  $r_d = .75$ , respectively; and the *K* Scale, where  $r_d = .81$  among males and  $r_d = .73$  among females. The remaining five scales show coefficients at a lower level.

DISCUSSION

In general these results are in line with the reputation of the various scales: those commonly thought of as "good" scales have emerged with high  $r_d$  coefficients, and those thought of as "poor" scales with lower ones. There may perhaps be some surprise that the coefficients shown by Scale 2 (*D*) are not higher, particularly among females, or that those shown by Scale 9 (*Ma*) are not lower. The outstanding characteristic of the results, however, is the enormous variation in the size of the  $r_d$  coefficients from scale to scale. Two of the scales show excellent dimensionality, three more are fairly satisfactory, but five of the scales show coefficients too low to warrant treating them as interval scales, at least among normal subjects. The bearing of these differences on the use and interpretation of the various scales will be briefly considered.

On the two most highly dimensional scales, Scale 7 (*Pt*) and Scale 1 (*His*), scale scores arrived at by the use of the standard keys reflect very closely the major systematic differences in the performances of normal subjects. On Scale 7, for example, the dimensionality coefficients of .96 and .91 indicate that scale scores account for 92% and 83% of the variance in the underlying dimensional positions or first factor scores of normal males and normal females, respectively. If two normal subjects obtain similar scores on these scales, the probability is very high that they have given similar performances. While such similarity does not in itself justify uniform inferences as to psychological meaning, it does provide a consistent behavioral basis for profile interpretations and correlational studies, including factor analysis. In this connection it should be noted that while Scales 1 and 7 are highly dimensional, they are also factorially complex. Comrey's analysis of the items of Scale 1 yielded eight identifiable factors, as did his analysis of Scale 7 (Comrey, 1957,

TABLE 1  
COEFFICIENTS OF DIMENSIONALITY OF MMPI  
SCALES IN TWO NORMAL SAMPLES

Scale	Coefficient of dimensionality, $r_d$		
	Male sample	Female sample	Difference
<i>K</i>	.81	.73	.08
1 ( <i>His</i> )	.93	.90	.03
2 ( <i>D</i> )	.67	.45	.22
3 ( <i>Hy</i> )	.11	.10	.01
4 ( <i>Pd</i> )	.36	.14	.22
5 ( <i>Mf</i> )	.61	.01	.60*
6 ( <i>Pa</i> )	.43	.41	.02
7 ( <i>Pt</i> )	.96	.91	.05
8 ( <i>Sc</i> )	.84	.79	.05
9 ( <i>Ma</i> )	.82	.75	.07

Note.— $N = 40$  in each sample.  
\*  $p < .01$ .

1958c). As may be expected from the high dimensionality of the scales, many of Comrey's factors are very small, contributing little more than random variation to overall scale scores. Eliminating these small factors would make only a slight improvement in the dimensionality of the scales; the chief gain would arise from making the scales shorter.

Somewhat less satisfactory, but still showing a moderate degree of dimensionality, are Scales 8 (*Sc*) and 9 (*Ma*) and also the *K* Scale. Over 50% of the variance in scores on each of these scales can be accounted for by a single dimension in the normal samples used in this study. While the relationship between score and performance here is less predictable than on the highly dimensional scales it is still high enough to justify treating these three scales as interval scales. As with the scales described above, factor analyses of the items of these show many small but identifiable factors. For Scale 8 Comrey and Marggraph (1958) identified 12 factors; for Scale 9 and the *K* Scale, Comrey (1958ab) identified 14 and 8, respectively. Because scale scores on these three scales are already so nearly dimensional, however, their dimensionality would be improved considerably by the elimination of a small number of cross-cutting items. Eliminating such items would not require any change in the character or interpretation of the scales. Rather it should serve to make existing interpretations more accurate.

In contrast to the foregoing are Scales 2-6, no one of which shows a dimensionality coefficient among normal subjects high enough to justify treatment as an interval scale. Since each of these five scales presents a somewhat different problem of use and interpretation, each will be considered separately.

Of the five, Scale 2, the Depression Scale, appears to have most in common with the more highly dimensional scales. While its dimensionality coefficients among normal subjects are low, they are relatively closer to an acceptable level than those of Scales 3-6. In two samples of hospital patients, the writer has obtained dimensionality coefficients for Scale 2 of .86 and .87, sufficiently high to make gross dimensional distinctions. To

permit consistent differentiation within the restricted range of talent represented by normal subjects, however, requires the elimination of many error items. The writer has developed a modified version of Scale 2 that appears to meet the strictest dimensional requirements, yielding coefficients of better than .90 in normal samples. A paper on this modified scale is in preparation.

Scale 3, the Hysteria Scale, on the other hand, shows no evidence of dimensionality among normal subjects. The two coefficients,  $r_d = .11$  for males and  $r_d = .10$  for females, indicate that scale scores account for only 1% of the variance in the *Ss'* positions on the major underlying dimension or factor. Two normal subjects who obtain the same score are no more likely to have given similar performances than are two subjects with widely different scores. Treating such a congeries as an interval scale makes very little sense, and it is not surprising that factorial studies of normal subjects show inconsistent results for this scale (e.g., Kassebaum, Couch, & Slater, 1959; Wheeler, Little, & Lehner, 1951). Making a dimension of Scale 3 would be an arbitrary procedure, for there is no dimensional nucleus or core that can be made more prominent. Yet obviously Scale 3 has proved very useful in profile interpretation and configural scoring (Welsh, 1952; Meehl & Dahlstrom, 1960). Indeed, the scale seems to show just the sort of nondimensional but consistent relationships that the configural approach was devised to tease out (Meehl, 1950). A given score on Scale 3, say a *T* score of 70, for example, is likely to arise from an entirely different set of items when a subject's Scale 2 score is low than when it is high. Presumably such grossly different performances can appropriately be given different interpretations. In short, the psychological significance of Scale 3 scores among normal subjects appears to be a matter of complex contextual interpretation, and the assumption of dimensionality is an error to be avoided.

Most of what has been said about Scale 3 is also appropriate to Scale 4, the Psychopathic Deviate Scale. In the original paper on this scale, McKinley and Hathaway



(1944) describe it as deliberately mixed in factor content and see its value as best illustrated in the diagnostic evaluation of the psychopathic personality in the clinic. Scale 4 has, however, also been reported as useful in making distinctions among normal subjects (e.g., Hovey, 1953). As with Scale 3, improving the dimensionality of this scale may actually impair its usefulness by reducing its validity in configural scoring and profile interpretation. This, however, is a question that is subject to test.

Scale 5, the Masculinity-Femininity Scale, has played a relatively unimportant role in the development and use of the MMPI (Hathaway, 1956). It is of interest here in good part because it is the only scale that shows a significant sex difference in dimensionality among normal subjects. While it is tempting to interpret this difference as supporting the hypothesis that femininity is not the mirror-image, or even the approximate mirror-image, of masculinity, no interpretation is justified at this stage. Among males, where  $r_d = .61$ , it should be possible to construct a highly dimensional scale from the items of Scale 5, but among females, where  $r_d = .01$ , such a scale could not be developed without changing the direction of scoring of many of the items.

The last of the scales showing low coefficients of dimensionality, Scale 6, the Paranoia Scale, is similar to Scale 5 in the sense that it seemed weak to its originators at the time of its publication and has played a relatively unimportant role in the subsequent use of the MMPI (Hathaway, 1956). Among normal subjects of both sexes the variance of scores is small, and it is not surprising that dimensionality coefficients of only .43 and .41 are found for male and female samples, respectively. An item analysis conducted by the writer indicates that the low dimensionality of Scale 6 is to be attributed in good part to subtle items that consistently draw more frequent endorsements from low than high scorers among both normal and hospital samples. Since Scale 6 has not distinguished itself by the accuracy with which it separates out persons with paranoid symptoms, there may be considerable merit in

improving the dimensionality of the scale by reversing the direction of scoring the most consistent of these subtle items.

## REFERENCES

- COMREY, A. L. A factor analysis of items on the MMPI Hypochondriasis Scale. *Educ. psychol. Measmt.*, 1957, 17, 568-577.
- COMREY, A. L. A factor analysis of items on the MMPI Hypomania Scale. *Educ. psychol. Measmt.*, 1958a, 18, 313-323.
- COMREY, A. L. A factor analysis of items on the K Scale of the MMPI. *Educ. psychol. Measmt.*, 1958b, 18, 633-639.
- COMREY, A. L. A factor analysis of items on the MMPI Psychasthenia Scale. *Educ. psychol. Measmt.*, 1958c, 18, 290-300.
- COMREY, A. L., & MARGGRAFF, W. M. A factor analysis of items on the MMPI Schizophrenia Scale. *Educ. psychol. Measmt.*, 1958, 18, 301-311.
- DEMPSEY, P. Liberalism-conservatism and party loyalty in the U. S. Senate. *J. soc. Psychol.*, 1962, 62, 159-170.
- DEMPSEY, P., & BAUMHOFF, M. The statistical use of artifact distributions to establish chronological sequence. *Amer. Antiq.*, 1963, 28, 496-509.
- GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer (Ed.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. Pp. 60-90.
- HATHAWAY, S. R. Scales 5 (Masculinity-Femininity), 6 (Paranoia), and 8 (Schizophrenia). In G. S. Welsh and W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 104-111.
- HOVEY, H. B. MMPI profiles and personality characteristics. *J. consult. Psychol.*, 1953, 17, 142-146.
- KASSEBAUM, G., COUCH, A., & SLATER, P. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, 23, 226-236.
- LAZARSFELD, P. F. The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer (Ed.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. Pp. 362-412.
- LOEVINGER, JANE. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, 61(4, Whole No. 285).
- LORD, F. M. A theory of test scores. *Psychometrical Monogr.*, 1952, No. 7.
- LUMSDEN, J. The construction of unidimensional tests. *Psychol. Bull.*, 1961, 58, 122-131.
- MCKINLEY, J. S., & HATHAWAY, S. R. The MMPI: V. Hysteria, hypomania and psychopathic deviate. *J. appl. Psychol.*, 1944, 28, 153-174.
- MEERL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, 14, 165-171.



- MEEHL, P. E., & DAHLSTROM, W. G. Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *J. consult. Psychol.*, 1960, 24, 375-387.
- WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, 16, 65-72.
- WELSH, G. S. Factor dimensions A and R. In G. S. Welsh and W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 264-281.
- WELSH, G. S., & DAHLSTROM, W. G. (Eds.). *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, 15, 134-141.

(Received February 28, 1963)

## DEFINITIONAL ASPECTS OF THE CRISIS CONCEPT

BERNARD L. BLOOM

*Harvard School of Public Health*<sup>1</sup>

8 expert judges reacted to a series of 14 miniature case histories of stress experiences which had been constructed so that it was possible to study the differential effects of 5 variables upon judgments of the presence or absence of "Crisis." The variables studied were (a) knowledge or lack of knowledge of a precipitating event, (b) rapidity of onset of reactions, (c) awareness or lack of awareness on the part of the hero of inner discomfort, (d) evidence of behavioral disorganization, and (e) rapidity of resolution. The results indicated that to these judges, a crisis was defined as an episode beginning with a known precipitating event followed by either no discernible reaction or by some reaction which required a month or somewhat longer to resolve. Within a group of episodes in which precipitating events were known, there appeared no way by which crises could be reliably judged on the basis of subsequent reactions. In view of the apparent lack of clarity of the crisis concept, further refinement is indicated before attempts to assess the effectiveness of crisis intervention are undertaken.

Among psychologically stressful experiences, those particular events referred to as crises are receiving increasing attention by mental health workers. Interest in the crisis has arisen partly because it is felt that the consequences of the management of such periods of stress often include significant and long-lasting changes in level of adequacy of mental functioning. Many workers who apply public health concepts to the field of mental health believe that good mental health is in large measure the result of a life history of successful crisis resolutions; and, therefore, by providing therapeutic intervention to people while they are in crisis, the incidence of subsequent mental disorder in these persons may be significantly reduced. Clinical studies of crises precipitated by specific events have appeared in recent years. Bereavement has been studied by Lindemann (1944); reactions to surgery have been studied by Janis (1958); and reactions to the birth of a premature child have been studied by Caplan (1960) and by Kaplan and Mason (1960). These studies are related to earlier investigations of combat reactions and responses to natural disasters, although it is only within the past several years that the concept of crisis has been given focal importance.

<sup>1</sup> Now with National Institute of Mental Health at the Denver Regional Office of the United States Public Health Service.

During the exploratory phase of the investigation of many new concepts, there is a danger that definitions can be too limiting. The boundary between what is included and what is not included in a particular concept might appropriately be kept fluid and deliberately vague so as not to restrict unduly the span of inquiry. Precise definitions of concepts are required, however, as one advances beyond initial exploration toward the testing of hypotheses which have been advanced involving the concepts. This phase appears to have been reached regarding the concept of crisis. Caplan (1961) has suggested, for example, that "during the period of upset of a crisis, a person is more susceptible to being influenced by others than at times of relative psychological equilibrium [p. 13]." Referring to this increased susceptibility, Caplan (1962) suggests that "from a preventive psychiatric point of view, this is a matter of supreme importance; because by deploying helping services to deal with individuals in crisis, a small amount of effort leads to a maximum amount of lasting response [p. 82]." Whatever else may be required to test this provocative hypothesis, relevant life events must be unambiguously sortable as either crises or not crises. The success of studies of attempts to intervene and influence the outcome of such life experiences will be partially dependent on the adequacy of this fundamental definition. The present investigation has been

designed to explore the nature of agreement with respect to the crisis concept among clinicians, highly skilled in both crisis theory and community mental health practice.

### METHOD

In the normal noncrisis state the existing repertoire of responses available to an individual is adequate to meet and solve problems as they occur. A crisis occurs when an individual finds himself unable to deal effectively with an emerging problem. Crises vary in severity, with mild crises shading into the usual ups and downs of problem solving behavior which characterize the ordinary functioning of an individual. Review of the relevant literature suggests that the crucial elements in the identification of the crisis-state appear to be (a) a stressful precipitating event of which the individual is aware; (b) significant subsequent rapid cognitive and affective disruption unusual for that particular individual; and (c) duration of the disruption of at least several days. With regard to these elements a number of questions may be raised concerning the behavior of clinicians in judging the presence or absence of a crisis. Is a reaction sequence viewed as a crisis if there is no awareness of a precipitating event either by the individual himself or by significant persons in his environment? How rapid must behavioral changes be in order to suspect that a crisis is present? Does a person need to have some awareness of inner discomfort or tension in order to be judged to be in crisis? Is a person considered to be in crisis if there are no external manifestations of disequilibrium? Does the rapidity of the resolution of the stressful experience have any bearing on whether the experience is judged as constituting a crisis?

The test instrument utilized in this investigation consisted of 14 brief (averaging approximately 175 words each) case histories. These fictional histories each described a single stressful event in which the central character was named Mr. Jones. Each case history was constructed to appear as realistic as possible, yet each contained information bearing on (a) the awareness or lack of awareness of a precipitating event; (b) the rapidity of the onset of the disruption; (c) presence or absence of internal discomfort; (d) external evidence of behavioral disruption on the part of the presumed victim of the crisis; and (e) rapidity of resolution of the conflict. Starting by dichotomizing each of these five components, 13 of the stories were so constructed that the five variables under study appeared in more or less random conjunction with each other. Thus, for example, Story B was composed in which a known precipitating event was followed by sudden onset of both internal discomfort and external manifestations of disequilibrium following which there was slow resolution of the conflict. Story M was composed in which there was gradual onset of externally visible disorganization, no awareness of internal tension, no known

precipitating event, and rapid resolution. The fourteenth story (Story N) involved a known precipitating event followed by no symptoms of any kind, neither internal discomfort nor behavioral disorganization. A group of professionally trained mental health workers served as judges during the construction of these stories, and a story was included in the final test instrument if there was 80% agreement by the judges as to the presence or absence of all five of the components intended to characterize each story. As examples of the fourteen stories, Stories B, M, and N are herewith presented.

### Story B

Mr. Jones slammed on the brakes but not in time to avoid hitting the boy who had dashed out in the street in front of his car. Before he could even open the door of the car, he felt nauseated and very frightened. He managed, almost blindly, to reach the front of the car but found himself unable to do anything to assist the moaning teen-ager who was badly cut and bleeding. When help arrived, Mr. Jones was in a dazed condition, unable to talk coherently about the accident but aware that he was tremendously disturbed. Fortunately, the boy's injuries seemed much more severe than they actually were, and he recovered with no permanent injuries. Although Mr. Jones was found not to be legally responsible for the accident—and the boy whom he hit fully admitted that it was not Mr. Jones' fault—it was months before Mr. Jones was able to talk about the accident and drive his car comfortably.

### Story M

Mr. Jones' peculiar behavior had begun very gradually. At first he seemed to become forgetful, then simply "off on another planet" most of the time. When Mrs. Jones would try to call it to his attention, she found him completely unaware of his actions. Then one day he just disappeared. There was simply no trace of him for nearly a week, when he was found by a United States Customs officer returning to New York on a flight from Europe. He was arguing with the officer about paying duty on a guitar he had bought in Spain, and he had attracted quite a crowd. He was very belligerent and the officer, suspecting he was not well, called the police—who in turn called an ambulance. In the ambulance Mr. Jones suddenly seemed to realize who and where he was and, after satisfying the medical authorities that he was well, was released. He was home again the next day, virtually back to normal.

### Story N

The day after Mrs. Jones was killed in a plane accident, Mr. Jones reported for work on time just as he had done more or less regularly for the past several years. The office staff were rather



TABLE 1  
ENUMERATION OF THE COMPONENTS OF THE TEST STORIES

Precipitating event			Onset of symptoms		Internal tension		Behavioral disorganization		Resolution of conflict	
Story	Known	Un-known	Sudden	Gradual	Recognized	Not recognized	Visible	Not visible	Resolution of conflict	
									Rapid	Slow
A	X		X		X		X		X	
B	X		X		X		X			X
C	X		X		X			X	X	
D	X		X		X			X		X
E	X		X			X	X		X	
F	X		X			X	X			X
G	X			X	X		X			X
H	X			X	X			X	X	
I	X			X		X	X		X	
J	X			X		X	X			X
K		X	X		X		X			X
L		X		X	X			X	X	
M		X		X		X	X		X	
N	X		None			X		X	Not relevant	

surprised to see him, particularly since Mrs. Jones' death had been so tragic—and they had been so much in love. The staff remarked to each other how well he was reacting to the tragedy and how no one could ever have guessed that so terrible a thing had happened so recently. Mr. Jones' boss was particularly understanding and tried to induce him to take time off, but Mr. Jones seemed willing to work and was as friendly and relaxed as ever. Weeks later Mr. Jones remarked that his wife's death had been a blow to him, but life had to go on. He said that he had never felt any serious discomfort or any real difficulty in carrying on his normal activities.

Enumeration of the components of the 14 test stories is presented in Table 1. As will be seen, there are considerably more stories in which a precipitating event is present than in which no precipitating event is known. Except for this variable, the stories distribute themselves nearly equally with regard to each of the components under study.

The test instrument was individually administered to eight experts in the field of crisis theory, all members of the staff of the Community Mental Health Program at the Harvard School of Public Health. Each story appeared on a separate sheet followed by the same two questions. The first asked, "On the basis of your understanding of the concept of 'crisis,' did this event constitute a crisis for Mr. Jones?" The judge was asked to check one of the three choices, namely, "yes," "no," or "do not know." The second question was, "In a sentence or two, why did you check this alternative?"

## RESULTS

The measure employed to determine the extent to which each event was judged to constitute a crisis was to divide the total number of "yes" judgments by the sum of the "yes" and "no" judgments on each story. This measure treats the "don't know" judgments as qualitatively different from both "yes" and "no," and a separate estimate of uncertainty was made by calculating the proportion of all judgments to each story which were in the "don't know" category. In only 5 of the 14 stories were the crisis judgments unanimous. In the case of the remaining 9 stories, "yes" percentages ranged from 20 to 86. In 4 of the stories, the "yes" percentages were between 40 and 63. "Don't know" percents ranged from 0 to 50. In only 5 stories were there no "don't know" judgments. The judges were thus clearly not unanimous in their judgments.

In the analysis of the judgments of crisis as a function of the five components under study, the set of 14 stories was repeatedly subdivided so that judgments to all stories containing each particular component could be contrasted with judgments to all stories not containing that same component. Analysis

of the relationship of these components to judgments of crisis is presented in Table 2. Of the five variables, the crisis judgment seems related to two. The judgment of crisis is made significantly more often when there is a known precipitating event than when the precipitating event is unknown—and is made significantly more often when there is slow resolution as contrasted with rapid resolution. Judgments of crisis are made with particular difficulty when the precipitating event is unknown. Under this circumstance there is a high level of uncertainty on the part of the judges. With respect to knowledge of the precipitating event, it should be pointed out that the six individuals who made a crisis judgment to Story N are unanimous in considering it a crisis—although in this story no reactions follow the event. Crisis judgments

appear to be unrelated to variations in the other three variables.

A rapid resolution was defined as one which took place within a week; a slow resolution was one which required between 1 and 2 months. In the case of this variable, episodes in which there is slow resolution are judged significantly more often to be crises than episodes in which the resolution is rapid. In view of the fact that episodes with known precipitating events tend to be judged as crises and episodes with rapid resolution tend not to be judged as crises, it might be anticipated that histories characterized by *both* known precipitating events *and* rapid resolution would be particularly ambiguous to the judges. Five stories share these two attributes, and "yes" percents to these particular stories range between 57 and 72,

TABLE 2  
CRISIS JUDGMENTS AS A FUNCTION OF THE EXPERIMENTAL VARIABLES

Experimental variable		No. of stories	Judgments					
			N	Yes	No	Dk	Yes%	Dk%
Precipitating event	Known	11	88	65	16	7	80	8
	Unknown	3	24	3	11	10	21	42
	Difference						59	34
	S.E. Diff.						11.7	10.5
	C.R.						5.0*	3.2*
Rapidity of onset	Sudden	7	56	36	13	7	73	13
	Gradual	6	48	26	14	8	65	17
	Difference						8	4
	S.E. Diff.						9.8	7.0
	C.R.						ns	ns
Internal tension	Recognized	8	64	36	18	10	67	16
	Not recognized	6	48	32	9	7	78	15
	Difference						11	1
	S.E. Diff.						9.1	6.9
	C.R.						ns	ns
Behavioral disorganization	Visible	9	72	43	18	11	70	15
	Not visible	5	40	25	9	6	74	15
	Difference						4	0
	S.E. Diff.						9.5	7.0
	C.R.						ns	ns
Speed of resolution	Rapid	7	56	27	20	9	57	16
	Slow	6	48	35	7	6	83	13
	Difference						26	3
	S.E. Diff.						9.3	6.9
	C.R.						2.8*	ns

\*  $p \leq 0.01$ .

indicating extensive disagreement among the judges.

The results of the analysis would seem to indicate that in practice, a crisis is defined primarily in terms of knowledge of a precipitating event and secondarily in terms of a slow resolution. Known precipitating events are generally judged to lead to crises if (a) there is no reaction or if (b) there is a reaction of any kind and resolution requires a month or more. The judges' comments suggest that situations in which the resolution is rapid are commonly viewed as episodes illustrating appropriate responses to reality situations. Reactions of any kind which appear when there is no known precipitating event are likely to be considered psychiatric disorders rather than crises. Since in practice intervention ordinarily takes place after the identification of some precipitating event which might lead to crisis but before resolution of any resultant stress, the data were reanalyzed in order to determine whether, within the group of 11 episodes in which a precipitating event was known, crisis judgments were related to the nature of subsequent reactions. Crisis judgments in the case of the histories in which there was a known precipitating event and sudden onset were contrasted with crisis judgments in which there was a known precipitating event and gradual onset. Similarly, comparison was made between judgments in instances of known precipitating events and recognized internal tension and judgments in instances of known precipitating events and unrecognized internal tension. The final analysis contrasted crisis judgments in cases of known precipitating events and visible behavioral disorganization with judgments in cases of known precipitating events and no visible behavioral disorganization. In no case was a significant difference found.

#### DISCUSSION

In order to test the efficacy of intervention at times of crisis, a sample of people in crisis must be identified. Following this identification, one could contrast outcome in a subgroup who had been exposed to inter-

vention procedures with another subgroup who had not been exposed to such procedures. Outcome in both subgroups could be compared with that in a noncrisis group. But whether or not a controlled study of this kind is undertaken, the identification of the crisis subgroup should be sufficiently unambiguous so that it is the *intervention* which is clearly the subject of study. Failure of intervention procedures should not be attributable to misdiagnosis of the crisis state. The single most contributory variable involved in the crisis judgment in the present sample appears to be the existence of a precipitating event. On this basis, one could simply define the crisis state as inevitably following certain specified events. This is an appealing operational solution to the task of definition, but judges do not maintain that crises inevitably follow certain events. If one does not define crisis solely by the existence of some event of the life of the person, actual practice would dictate that within a group in which some specific event occurs, one should be able to distinguish between those people for whom the event results in a crisis state and those people who seem not to be in crisis as a consequence of the event. The present findings suggest that this kind of discrimination may not now be possible. Until further refinement of the crisis concept is undertaken, assessment of the effectiveness of intervention efforts will be difficult.

#### REFERENCES

- CAPLAN, G. Patterns of parental response to the crisis of premature birth. *Psychiatry*, 1960, 23, 365-374.
- CAPLAN, G. *Prevention of mental disorders in children*. New York: Basic Books, 1961.
- CAPLAN, G. *Manual for psychiatrists participating in the peace corps program*. Medical Program Division, Peace Corps. Washington, D. C., 1962.
- JANIS, I. L. *Psychological stress*. New York: Wiley & Sons, 1958.
- KAPLAN, D. M., & MASON, E. A. Maternal reactions to premature birth viewed as an acute emotional disorder. *Amer. J. Orthopsychiat.*, 1960, 30, 539-552.
- LINDEMANN, E. Symptomatology and management of acute grief. *Amer. J. Psychiat.*, 1944, 101, 141-148.

(Received March 15, 1963)



## THE EDWARDS SOCIAL DESIRABILITY SCALE AS A SHORT FORM OF THE MMPI

MAURICE O. ELVEKROG<sup>1</sup> AND NORRIS D. VESTRE<sup>2</sup>

*Veterans Administration Hospital, St. Cloud, Minnesota*

This study evaluates the accuracy of prediction of MMPI scale scores from the Edwards' SD scale. SD-predicted MMPI scores were obtained according to a procedure presented by Edwards and were compared with actual scores for 1 normal and 2 psychiatric samples. Analysis of the relationship between predicted and obtained scores via product-moment correlations, testing of the significance of the differences between the means, and a tabulation of the predictive errors indicates that the predictive efficiency of the Edwards' procedure shows striking variability with different MMPI scales and, to a lesser extent, with different population samples. In general, the results failed to support the use of the Edwards SD scale as a "short MMPI."

The tremendous amount of work that has recently been done on "response styles" in psychological tests has culminated in the claim by Edwards (1962) that at least for the Minnesota Multiphasic Personality Inventory (MMPI), the variance of the test scales is so largely a function of social desirability that an individual's MMPI scores can be predicted simply on the basis of his Social Desirability (SD) score, and that, in fact, the SD scale is "a short form of the MMPI" (Edwards & Walker, 1961). While there are serious methodological problems involved in the evidence offered for this claim, the present study is confined to the question of whether the procedure suggested by Edwards is of any practical utility as a short form of the MMPI. Holloway and Gocka (1962) have published a table based on the Edwards prediction equation. In the present study SD-predicted MMPI scores were obtained from this table which gives the predicted values for the usual clinical and validity scales of the MMPI for all possible SD scale scores (0 to 39).

### METHOD

#### *Sample*

Three different samples were studied since Edwards and Walker (1961) assert that prediction may be

more accurate for some groups (with high SD scores) than for others. Sample A was composed of 76 chronic schizophrenics who had been hospitalized continuously in a Veterans Administration neuropsychiatric hospital for approximately 4 years. Their mean age was 39 years. Sample B consisted of 102 consecutive admissions to the same hospital. These had a mean age of 38 years. Fifty-eight percent had a psychotic diagnosis (most often schizophrenia); 30% had a neurotic diagnosis; the remainder had a personality trait disturbance, character disorder, or brain syndrome as a diagnosis. The admissions group, being more acutely disturbed, was expected to have more elevated MMPI profiles and lower SD scores than the chronic group. Group C was composed of 78 student nurses affiliated with this hospital. This group, whose mean age was 20 years, was expected to have higher SD scores than the patient groups and to resemble more closely the college students which have been the subjects of many of the earlier studies on this topic.

#### *Procedure*

The MMPI (Booklet Form) was administered in the usual way and scored for the Edwards SD scale in addition to the usual clinical and validity scales. For each subject, the Holloway and Gocka table was entered with the individual's SD score and the predicted MMPI scale scores were read. These SD-predicted scores were *K*-corrected, where required, and converted to *T* scores for comparison with those obtained from the scoring of the actual MMPI records.

Three analyses were done: (a) Product-moment correlations between predicted and observed scores were compared and tested for significance. (b) The means of the differences between observed and predicted scores were evaluated by means of a *t* test. (c) A table of cumulative frequency percentages of absolute differences between predicted and observed scores was made. This may be the most relevant

<sup>1</sup> Now at Veterans Administration Hospital, Minneapolis, Minnesota.

<sup>2</sup> Now at University of Minnesota.

TABLE 1

CORRELATIONS BETWEEN OBSERVED AND SD-PREDICTED MMPI SCALE *T* SCORES

Scale	Chronics ( <i>N</i> = 76)	Admissions ( <i>N</i> = 102)	Nurses ( <i>N</i> = 78)
<i>L</i>	-.30**	-.10	-.15
<i>F</i>	.51**	.50**	.46**
<i>K</i>	.74**	.72**	.65**
<i>Hs</i>	.25*	.66**	-.01
<i>D</i>	.47**	.57**	.50**
<i>Hy</i>	.07	.32**	.10
<i>Pd</i>	.00	.14	-.21
<i>Mf</i>	-.27*	-.02	.04
<i>Pa</i>	.69**	.58**	.19
<i>Pt</i>	.71**	.48**	.37**
<i>Sc</i>	.82**	.59**	-.13
<i>Ma</i>	.46**	.20*	-.06
<i>Si</i>	.73**	.71**	.65**

\*  $p < .05$ .\*\*  $p \leq .01$ .

consideration for clinical purposes, as from it can easily be determined the frequency with which predictive errors of specified magnitudes existed in each sample.

## RESULTS

Product-moment correlations between *SD*-predicted and observed MMPI scores are presented in Table 1. As can be seen, there are considerable differences in level of correlations obtained for different scales and different samples. Consistently high correlations are produced in the case of the *K* and *Si* scales and, to a lesser extent, with the *F*, *D*, *Pa*, *Pt* and *Sc* scales. It is worth pointing out, however, that of the 39 coefficients in Table 1, nine are negative (only two of these nine are significant); also, six more are not significantly different from zero. All of the remaining coefficients (24 in number) are positive and significant. In the case of the *L*, *Hy*, *Pd*, and *Mf* scales, in particular, prediction is quite inaccurate. The correlations associated with these scales are either uniformly, or almost uniformly, negative or nonsignificant. The correlations for the student nurse sample are lower, on the whole, than for the other two samples. The average *r* values for the chronics, admissions, and nurses are .38, .41, and .18, respectively. It will be noted that

there are a number of instances wherein the student nurse sample stands out by virtue of disturbing the pattern of coefficients. For example, for the two psychiatric samples positive and significant coefficients obtain for the *Hs*, *Sc*, and *Ma* scales while for the student nurses low negative correlations were found on these three scales.

Table 2 indicates that for most of the scales there are systematic errors in prediction that are statistically reliable. Predictive accuracy could, therefore, be improved by the addition or subtraction of a constant appropriate to the scale in question, although this improvement would be limited by the size of the correlations reported in Table 1. There seems to be consistency between these three samples on whether a scale is overpredicted or underpredicted, but there is one obvious exception to this rule. This exception is the *D* scale which is underpredicted for the two psychiatric samples and overpredicted for the nurses.

Another procedure which would seem to offer meaningful data by which to evaluate the accuracy of *SD*-predicted MMPI scale scores is a tabulation of absolute differences between observed and *SD*-predicted *T* scores. In Tables 3, 4, and 5 are presented cumulative frequency percentages of predictive errors

TABLE 2

ANALYSIS OF MMPI *T* SCORE MEAN DIFFERENCES (OBSERVED MINUS *SD*-PREDICTED)

Scale	Chronics ( <i>N</i> = 76)	Admissions ( <i>N</i> = 102)	Nurses ( <i>N</i> = 78)
<i>L</i>	-2.99*	-7.56**	-11.50**
<i>F</i>	-10.88**	-27.45**	-18.73**
<i>K</i>	-2.33**	-4.93**	-5.78**
<i>Hs</i>	-4.03*	-8.86**	-5.22**
<i>D</i>	2.08	3.38*	-3.56**
<i>Hy</i>	-4.50**	-3.32**	-1.76*
<i>Pd</i>	5.11*	1.37	2.04
<i>Mf</i>	-16.41**	-11.97**	-11.44**
<i>Pa</i>	-6.21**	-11.47**	-12.38**
<i>Pt</i>	-3.69**	-2.94*	-1.01
<i>Sc</i>	-10.14**	-20.77**	-12.91**
<i>Ma</i>	-8.89**	-10.10**	-9.88**
<i>Si</i>	3.21**	4.34**	6.73**

Note.—Where mean difference is positive, obtained scores are higher; where it is minus, *SD*-predicted scores are higher.

\*  $p \leq .05$ .\*\*  $p \leq .01$ .

TABLE 3

CUMULATIVE FREQUENCY PERCENTAGES INDICATING DIFFERENCES BETWEEN OBSERVED  
AND PREDICTED MMPI *T* SCORES FOR CHRONIC SCHIZOPHRENIC SAMPLE

(*N* = 76)

Scale	Absolute differences between observed and predicted <i>T</i> score							
	<5	<10	<15	<20	<25	<30	<35	<40
<i>L</i>	42	55	83	87	97	100		
<i>F</i>	21	47	61	76	91	93	97	99
<i>K</i>	43	80	91	100				
<i>Hs</i>	26	58	71	79	83	89	93	93
<i>D</i>	14	39	68	83	93	93	96	97
<i>Hy</i>	42	62	78	84	91	93	95	97
<i>Pd</i>	25	39	54	70	78	87	93	95
<i>Mf</i>	9	20	42	58	82	92	97	100
<i>Pa</i>	24	51	72	88	96	99	100	
<i>Pt</i>	37	71	83	92	97	99	100	
<i>Sc</i>	24	39	63	82	88	93	97	99
<i>Ma</i>	22	53	67	83	92	96	100	
<i>Si</i>	42	80	91	97	100			

for the chronic schizophrenic, psychiatric hospital admissions, and student nurse samples, respectively. Although this type of analysis admittedly lacks the elegance of a statistical test, it may be argued that it is quite relevant for determining the practical utility of a "short MMPI."

A few aspects of Tables 3, 4, and 5 should be noted. The scale which is most accurately predicted over all three samples is *K*. For the total *N* of 256 (76 + 102 + 78) there are only three instances where *K* is missed by more than 19 *T* score points and for over 75% of these subjects the prediction error is not

TABLE 4

CUMULATIVE FREQUENCY PERCENTAGES INDICATING DIFFERENCES BETWEEN OBSERVED  
AND PREDICTED MMPI *T* SCORES FOR NP ADMISSIONS SAMPLE

(*N* = 102)

Scale	Absolute differences between observed and predicted <i>T</i> score							
	<5	<10	<15	<20	<25	<30	<35	<40
<i>L</i>	25	35	69	83	99	100		
<i>F</i>	12	23	27	31	42	47	61	70
<i>K</i>	39	74	87	97	100			
<i>Hs</i>	16	34	61	72	81	87	96	97
<i>D</i>	22	46	70	83	89	93	95	98
<i>Hy</i>	27	60	75	89	95	96	98	99
<i>Pd</i>	20	42	61	70	80	88	93	96
<i>Mf</i>	11	31	57	76	88	94	99	100
<i>Pa</i>	27	47	60	72	88	95	96	98
<i>Pt</i>	27	50	78	89	95	98	100	
<i>Sc</i>	12	22	33	47	53	64	75	82
<i>Ma</i>	16	38	58	76	87	94	98	99
<i>Si</i>	32	68	91	98	100			



TABLE 5

CUMULATIVE FREQUENCY PERCENTAGES INDICATING DIFFERENCES BETWEEN OBSERVED  
AND PREDICTED MMPI *T* SCORES FOR STUDENT NURSES SAMPLE

(*N* = 78)

Scale	Absolute differences between observed and predicted <i>T</i> score							
	<5	<10	<15	<20	<25	<30	<35	<40
<i>L</i>	13	27	73	85	100			
<i>F</i>	6	17	44	58	73	86	90	92
<i>K</i>	41	73	90	100				
<i>Hs</i>	40	65	86	99	100			
<i>D</i>	51	77	90	97	100			
<i>Hy</i>	45	79	94	100				
<i>Pd</i>	28	65	77	92	96	99	100	
<i>Mf</i>	19	37	67	83	94	99	100	
<i>Pa</i>	17	45	62	81	95	97	97	100
<i>Pt</i>	44	82	97	100				
<i>Sc</i>	23	41	55	72	83	94	97	97
<i>Ma</i>	24	47	65	82	90	95	97	100
<i>Si</i>	29	69	90	97	100			

greater than 9 *T* score points. Most of the other scales are not as accurately predicted. In particular, it might be noted that the predictions for *F* and *Sc* are quite inaccurate. In the case of *Sc*, approximately one-third of the total *N* of 256 are missed by over 19 *T* score points and approximately two-thirds are missed by over 9 *T* score points. Admittedly, the evaluation of an "acceptable" amount of error for practical purposes is a matter of judgment, but it would certainly appear that the amount of error involved in the prediction of *Sc*, *F*, and certain other scales could not be countenanced.

### DISCUSSION

The various analyses comparing *SD*-predicted and observed MMPI scale scores for the three samples indicates considerable variability in predictive efficiency depending upon scale and, to a lesser extent, on sample. Most of the observed correlations are indicative of a statistically reliable relationship between predicted and observed scores, but for some scales the relationship is extremely low and, for others, even negative. A statistically reliable difference between the means of ob-

tained versus predicted scores holds for nearly all the MMPI scales. A table of predicted *T* score errors of various magnitudes indicates that these errors are indeed sizable for some scales and, thus, that the Edwards prediction procedure could hardly serve for purposes of practical prediction. This does not vitiate the theoretical significance of the demonstrated relationships between the *SD* scale and many of the MMPI scales.

Results for the student nurse sample differ in a number of respects from results for the patient samples. This difference appears to be largely a function of a restriction of range in MMPI scores for the nurse sample. Thus, correlations obtained tend to be lower for this group. It also appears that the Holloway-Gocka table tends to give predictions within a restricted range and when a sample contains profiles almost all within the normal range, there is not the possibility of as widely erroneous predictions as might occur in patients groups having more deviant scores. It is clear also that some differences in results occur because the scores of the nurses sample tend to cluster at a different point on the scales than the patient scores do; this is exemplified in the case cited above where the *D*

scale is underpredicted for the patient samples and overpredicted for the student nurse sample.

## REFERENCES

EDWARDS, A. L. Social desirability and expected means on MMPI scales. *Educ. psychol. Measmt.*, 1962, 22, 71-76.

EDWARDS, A. L., & WALKER, J. N. A short form of the MMPI: The SD scale. *Psychol. Rep.*, 1961, 8, 485-486.

HOLLOWAY, H., & GOCKA, E. A table for obtaining expected scores of MMPI scales from a social desirability measure. *Newsltr. for Res. in Psychol.*, (Veterans Administration Center, Kecoughton, Virginia) 1962, 4, 66-69.

(Received March 25, 1963)

## AN UNDERLYING VARIABLE IN THE CLINICAL EVALUATION OF DRAWINGS OF HUMAN FIGURES

BERNARD I. LEVY<sup>1</sup>

*The George Washington University*

JAMES V. LOMAX, JR.<sup>2</sup>

*The George Washington University*

AND RAPHAEL MINSKY

*University of Maryland*

Artistic excellence correlates with evaluations of adjustment made by clinicians from drawings of human figures. In this study artistic excellence is defined, in part, as the accuracy of proportional relationships among parts of the body. 6 comparable sets of drawings varying along the dimension of proportional accuracy and 6 attributes inferred from drawings were selected. 36 clinical psychologists ranked drawings in accordance with the attributes. The results suggest that adjustment, artistic merit, dependency, intelligence, and sexual difficulty are all strongly related to proportional accuracy. The attribute of aggression is not. It is proposed that proportional accuracy rather than creative artistic values be considered as an underlying dimension in the clinical judgment of human figure drawings.

Whitmyre (1953) argues cogently that the artistic excellence of drawings of human figures is a very powerful influence upon the clinician's evaluation of emotional adjustment. Indeed one is left with the impression, after studying his results, that artistic excellence is perhaps all that is used as a basis of judgment. Sherman (1958) reported a study similar to Whitmyre's with essentially the same results and conclusions. The two studies help strengthen each other to the point where the present authors feel the combined results are an important assault upon the clinical use of figure drawings. Yet as conclusive as the two studies are, questions still remain. Neither Whitmyre nor Sherman attempts a definition of artistic excellence. They are wise. Furthermore both investigators limit themselves to the one attribute of psychological adjustment. In this study we approach both problems by proposing an operational definition of an aspect of artistic excellence and by extending the categories to be judged.

To avoid sinking into the morass of defin-

ing artistic excellence, we have selected one obvious element of the concept which we feel is crucial in the judgment of figure drawings—proportional accuracy. The morphology of the human figure contains relatively stable relationships among parts. The graphic artist, for example, may be accustomed to conceiving of and drawing the male figure as seven and a half head-lengths tall; shoulders are two head-lengths wide, arms are three head-lengths long, etc. (Bridgeman, 1928). The length of the head provides an easy unit for statement of ratios among body parts. It is to the accuracy of these ratios that clinicians and artists respond, we hypothesize, when they judge the drawings of human figures.

But does the clinician respond to proportional accuracy as he judges a wide variety of personality characteristics? The work of both Whitmyre and Sherman suggests the possibility of an affirmative answer at least for adjustment and artistic talent. Two studies by Albee and Hamlin (1949, 1950) also lend support to a "global judgment" concept of the adjustment dimension. In work with children's drawings, Coltharp (1961) found clinicians able to predict intelligence but unable to predict either sociability or emotional

<sup>1</sup> This work was done when the author was at the District of Columbia General Hospital and Georgetown University.

<sup>2</sup> Now at Saint Elizabeth's Hospital.



adjustment. Based upon the authors' own experience and the foregoing studies, three characteristics to be judged were isolated—*adjustment*, *artistic merit*, and *intelligence*. A count of traits from Machover's (1949) standard work suggested three other characteristics—*dependency*, *aggression*, and *sexual difficulties*. Having selected six characteristics which seem possible to judge from drawings of human figures we hypothesize that proportional accuracy has a strong influence upon clinicians' judgments of both general and specific characteristics of personality.

#### PROCEDURE

Thirty-six clinical psychologists were asked to rank order separate sets of drawings for each of the six selected characteristics. Every drawing had a known proportional accuracy score. The development of six comparable sets of nine drawings each proceeded as follows:

1. Three hundred and sixty drawings were available from a study done by Minsky and Levy (1961). Each drawing had been scored along seven dimensions of proportional accuracy as described in a manual by Levy and Minsky (1962). The seven scores were added together for an overall estimate of proportional accuracy after each distribution had been normalized and placed on the same scale by using the T Scale conversion.

2. The distribution of overall proportional accuracy scores was plotted. Using the distribution as a guide, 54 drawings were selected using the following criteria: (a) Six drawings had to be found at each of nine different levels of accuracy. The scores of each of the six drawings had to be identical or as homogeneous as possible. (b) Each of the nine groups of drawings had to be selected to cover the widest possible range of accuracy and yet have reasonably equal units between them. (c) None of the drawings selected could be unclothed, have missing limbs or have sexual organs showing through clothing. This latter criterion was used to avoid stimuli so salient and misleading to the clinician that their good sense might be abandoned.

3. After the six sets of nine drawings were selected an analysis of variance of the overall accuracy scores resulted in an  $F$  of .01 between sets. As far as homogeneity of scores within sets is concerned therefore, they are comparable.

4. Looking at the array of selected drawings however, their qualitative heterogeneity became obvious. Consequently 10 judges were asked to rank the six drawings at each accuracy level according to how good the drawing was. The averages of these ranks were taken and drawings assigned to sets accordingly. Thus Set I had all nine of those drawings ranked best, Set II had those ranked

next best and so on with Set VI having all of the nine drawings ranked lowest by judges. An analysis of variance of the accuracy scores of this new arrangement of drawings resulted in an  $F$  of .01 between sets.

5. Completed the development of the stimuli. Six sets of nine drawings were selected each with a wide range of proportional accuracy yet each with comparable levels of accuracy represented. In addition, an attempt was made to make each set qualitatively homogeneous.

Thirty-six subjects were selected to judge the drawings. Each subject is a clinical psychologist with experience in the interpretation of drawings of human figures. The mean of their experience is 6.7 years. There were 20 males and 16 females in the sample of judges. Judges were asked to rank order each set of drawings according to a different characteristic. A  $6 \times 6$  Latin square provided the presentation order of the six sets of drawings; the six traits were randomized and then assigned to the sets in the same order for each of the first block of six judges. The Latin square and assigned traits were then inverted for the second group of six judges, then used in the original order for the third group of judges and so on. This procedure provided a random and counterbalanced order of presentation. Finally the nine drawings in each set were randomized for presentation to the judges; this random order was also reversed for every other judge. The directions to judges were:

I am going to ask you to participate in an experiment on the clinical evaluation of drawings of human figures. In each of these six folders are nine drawings, each of which was done by a different person. I would like you to rank order the drawings in each folder according to the one of six dimensions which I will describe later. Thus, there will only be six rankings. Indicate your rank order by arranging the drawings in sequence. Brief descriptions of the end points will be provided. In cases of a tie among drawings, treat the ties as paired-comparison judgments.

Two cards containing brief descriptions of the "good" and "bad" extremes of each trait marked the end points of the ranks. At the completion of every ranking, the overall proportional accuracy score for each drawing was entered in the rank position given the drawing by the judge.

#### RESULTS

Analysis of variance is used: one category of analysis, judged assigned ranks, provides special problems of experimental dependence and correlation across ranks since a single judge ranked the nine drawings in each set. Consequently a conservative model for analysis is required. Edwards' (1950) model for

TABLE 1

VARIANCE RATIOS FOR TWO SOURCES OF VARIANCE AND EACH CHARACTERISTIC JUDGED

Characteristics	Sources	
	Ranks ( <i>df</i> = 8)	Interaction ranks × sets ( <i>df</i> = 40)
Adjustment	20.70**	1.56*
Aggression	.43	1.38
Artistic merit	25.20**	3.04**
Dependency	9.36**	.69
Intelligence	38.64**	2.97**
Sexual difficulty	2.90*	1.57*

\*  $p \leq .05$ .\*\*  $p \leq .01$ .

repeated measurements is used in this study. Rank is the source of variance of interest: the interaction, sets by ranks, is the most conservative error term for testing that effect and consequently is the one used.

Table 1 contains a summary of the analyses. These results are dramatic. Proportional accuracy is apparently a compelling cue utilized by clinicians in their judgments of general and specific personality characteristics. The only exception to the foregoing conclusion is the characteristic of aggression.

The results are further elaborated in Figure 1. The mean overall proportional accuracy score for each rank is computed by combining all scores across sets and dividing by the number of judges. Thus each mean is based upon 36 rankings. The mean accuracy scores for each of the six characteristics is

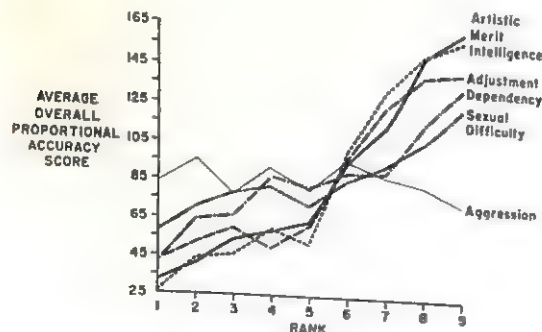


FIG. 1. The relationship between proportional accuracy and the rank assigned human figure drawings by clinical psychologists.

TABLE 2

CORRELATIONS BETWEEN PROPORTIONAL ACCURACY AND AVERAGE RANK FOR EACH CHARACTERISTIC JUDGED

(N = 54)

Characteristics	Correlations
Adjustment	.76*
Aggression	-.06
Artistic merit	.85*
Dependency	.63*
Intelligence	.84*
Sexual difficulty	.52*

\*  $p \leq .01$ .

plotted against rank. Every point, therefore, is the mean of the overall proportional accuracy score of drawings placed in that rank. In general, the curves for the five significant characteristics rise slowly until the middle rank and then accelerate sharply. This result is an artifact of the selection of drawings. Although attempts were made to keep distance between levels of accuracy constant, in fact the distances between levels became larger as overall accuracy decreased. The shape of the curves in Figure 1 when considered with the progressive variation in size of intervals between drawings suggests that judges make fine discriminations less easily than coarse ones.

Table 2 contains the results of a correlational analysis. Each of the 54 drawings is ranked by 6 judges for each characteristic. An average of the 6 ranks is related to the overall proportional accuracy score to provide a more rigorous statement of the relationships implied in Table 1 and Figure 1.

TABLE 3

AVERAGE INTRACLAS RELIABILITY COEFFICIENTS FOR EACH CHARACTERISTIC JUDGED AND FOR EACH SET

Characteristics	Sets					
	I	II	III	IV	V	VI
Adjustment	.65 <sup>a</sup>	.68	.75	.66	.85	.81
Aggression	.07	— <sup>b</sup>	.22	—	.26	.22
Artistic merit	.93	.92	.94	.92	.87	.76
Dependency	.10	.54	—	.51	.20	.45
Intelligence	.93	.92	.93	.90	.91	.84
Sexual difficulty	.36	.78	.25	—	.24	.31

<sup>a</sup> Based upon 6 judges and 9 drawings.<sup>b</sup> No relationship, error term greater than main effect.



Although reasonable degrees of interjudge reliability and reliability of means can be inferred from the results of the analyses of variance reported in Table 1, a more exact opportunity to test these reliabilities is described by Ebel (1951). The data easily lends itself to such analyses. Table 3 contains the results of this exploration of the stability of the mean ranks. The analyses are based upon groups of 6 judges each of whom ranked the same set of drawings for the same characteristic. Ranks rather than overall proportional accuracy scores are used.

### DISCUSSION

The results reported above seem to support the initial hypothesis in the main. Whether in or out of awareness, the clinician is responding to the proportional accuracy of the human figures he interprets. That proportional accuracy is basic to judgments of specific as well as general characteristics is not as easily documented. In the first instance, the attribute of aggression has no relationship to proportional accuracy. But the results of the intraclass reliability computations strongly suggest that judgments of this attribute are not reliable. For example in Sets II and IV, the variability of rankings is so extensive that it transcends any implication of central tendency for drawings. The four remaining coefficients are quite low. It is possible that aggression is too general a characteristic. Fractionating it into extropunitiveness and intropunitiveness might have yielded more reliable judgments if not a relationship to proportional accuracy.

Of the remaining five characteristics judged, *intelligence*, *artistic merit*, and *adjustment* seem, by virtue of the degree of their relationship to proportional accuracy and reliability, to be rather clearly associated with the experimental variable. Although the correlations between proportional accuracy and *dependency* and *sexual difficulty* are highly significant and reasonably high, their intraclass reliabilities suggest some limitations upon generalization. Table 3 does not reveal a systematic pattern of reliabilities by sets. It is as though a "groups of judges  $\times$  sets  $\times$  characteristic" interaction is at work. Some groups

of judges seem to respond differently to some sets of drawings and, in addition, respond differently depending upon the characteristic they are judging. It may be that central personality dimensions such as dependency and sexual difficulty are still too diffuse as concepts to enable reliable judgment. Perhaps too, sets with greater intervals of proportional accuracy between drawings might have resulted in more stable judgments.

A comment about artistic value as an explanatory variable in figure drawing judgments is appropriate here. If one considers the highly personal conception of the human figure revealed in paintings by contemporary artists such as George Cohen, Nathan Oliveira, and Jacob Landau to name a few, the notion of esthetic merit in its creative sense ought to be abandoned as an explanation of clinicians' judgments of drawings of human figures. Paintings by these artists can be found in the Museum of Modern Art exhibition catalogue, *Recent Painting USA: The Figure* (1962). This "new" image of the figure gives it primacy in the painting, emphasizes it and minimizes context. The figures are often distorted, featureless, and unclothed. They are painted with assertive strokes and usually without regard to anatomical precision. Human figures drawn in the spirit of these artists might easily be called "sick" by a clinician. This study suggests that one might better substitute proportional accuracy for esthetic merit as an underlying variable in clinical judgment of drawings. Indeed the .85 correlation coefficient between the two dimensions suggests that they are close to being interchangeable.

Although the results of this study seem conclusive, caution should be observed. Proportional accuracy is only one aspect of representational accuracy. Within the latter domain, variables such as symmetry, anatomical detail, clothing detail, posture, facial expression, etc. are all akin to proportion. However, since one expects all representational variables to be highly intercorrelated, proportional accuracy is probably as reasonable a representative of the domain as one can find, particularly because of the ease with which it can be measured.



## REFERENCES

- ALBEE, G. W., & HAMLIN, R. M. An investigation of the reliability and validity of judgments inferred from drawings. *J. clin. Psychol.*, 1949, 5, 389-391.
- ALBEE, G. W., & HAMLIN, R. M. Judgment of adjustment from drawings: The applicability of rating scale methods. *J. clin. Psychol.*, 1950, 6, 363-365.
- BRIDGEMAN, G. B. *Bridgeman's life drawing*. (3rd ed.) New York: Edward C. Bridgeman, 1928.
- COLTHARP, FRANCES G. Clinical judgment and the DAP test. *J. consult. Psychol.*, 1961, 25, 43-45.
- EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407-424.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart & Company, 1950.
- JUNIOR COUNCIL OF THE MUSEUM OF MODERN ART. *Recent painting USA: The figure*. New York: Doubleday, 1962.
- LEVY, B. I., & MINSKY, R. *Manual for the assessment of the accuracy of drawings of human figures*. Washington: B. I. Levy, 1962.
- MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
- LEVY, B. I., & MINSKY, R. Intelligence and adjustment in relation to the proportional accuracy of drawings of human figures. Unpublished study, University of Maryland, 1961.
- SHERMAN, L. J. The influence of artistic quality on judgments of patient and non-patient status from human figure drawings. *J. proj. Tech.*, 1958, 22, 338-340.
- WHITMYRE, J. W. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. *J. consult. Psychol.*, 1953, 17, 421-424.

(Received March 28, 1963)

# THE RELATIONSHIP BETWEEN NEUROTICISM, STRESS, AND QUALITATIVE PORTEUS MAZE PERFORMANCE<sup>1</sup>

ARISTOTLE A. ANTHONY<sup>2</sup>

*Teachers College, Columbia University*

The results of this experiment are the first empirical demonstration of an existing relationship between anxiety and Q score. Specifically, this research tested the following hypotheses: (a) the Q score of neurotic adolescents is higher than the Q score of normal adolescents ( $p < .001$ ), (b) the Q scores of both normal and neurotic adolescents are higher under interfering than under noninterfering test conditions ( $p < .01$ ), and (c) the difference in Q scores between interfering and noninterfering test conditions is greater for neurotic adolescents than for normal adolescents ( $p < .001$ ). For the experiment, 30 neurotic and 30 normal adolescents were tested, first under noninterfering test conditions, and then under interfering test conditions. The 2 groups were equated for age, intelligence, and parental socioeconomic level.

The purpose of this study was to investigate the relationship between anxiety and qualitative aspects of performance on the Porteus Maze test. This relationship was investigated by comparing the qualitative Porteus Maze performance, operationally defined by the Q score, of normal and neurotic adolescents under interfering and noninterfering test conditions.

The Q score was developed by Porteus in response to his observation that qualitative differences in maze performance occur. Two individuals may earn the same quantitative score or mental age, but the style or quality of their performance may differ widely. In 1942, Porteus and Honzik began a systematic study of qualitative errors in execution which were characteristic of the responses of delinquents. These responses became known as Q score. The characteristics of performance considered in the Q score (Porteus, 1950, pp. 161-162) are: (a) an error in the first third of the maze; (b) an error in the last third of the maze; (c) a cut corner; (d) a crossed line; (e) a lifted pencil at any point except the exit; (f) a wavy line; (g) a wrong direc-

tion; (h) each qualitative error in either the 5 or the 6 year maze.

Porteus (1942, p. 20; 1945, p. 101) found that the Q score significantly differentiated delinquents from nondelinquents. His results were confirmed in other studies by Wright (1944), and by Docter and Winder (1954). Thus, the Q score appears to be consistently successful in distinguishing between delinquent and nondelinquent samples.

Although Porteus developed the Q score with special reference to the behavior of delinquents, clinical observations suggested that an evaluation of qualitative aspects of performance on the Porteus Mazes may be a particularly sensitive measure of anxiety. For example, a person who is seen as anxious from a clinical point of view sometimes shows considerable muscular tension, and in extreme cases, observable muscle tremors. Thus, while the anxious person may successfully traverse the maze, smooth motor coordination during the performance may be disturbed by his muscular tension, and this disturbance would be manifested by wavy lines, cut corners, or crossed lines, the kinds of errors or behavior characteristics considered in Porteus' qualitative scoring. Another observation sometimes noted in clinical administration of the maze tests is the anxious person's perception of the maze as a "set of traps." That is, if a wrong turn is taken during the maze tracing, the result is entrance into a blind alley; the subject is "caught," or "trapped," and the

<sup>1</sup> This paper is based upon a doctoral dissertation at Teachers College, Columbia University, Department of Psychological Foundations and Services. Thanks are due to Joel R. Davitz, Chairman of the dissertation committee, and Paul E. Eiserer and Millie Almy, members of the dissertation committee, for their invaluable assistance and guidance.

<sup>2</sup> Now at the Ridgewood Public Schools, Ridgewood, New Jersey.

anxious person's tension tends to be increased. Thus, he may become impulsive, try to escape from the situation in one way or another, and in his impulsive efforts to escape, make errors in the relatively simpler mazes, or at the beginning or near the end of more complex mazes.

This study was designed to test the general hypothesis that there is a positive relationship between anxiety and qualitative aspects of performance on the Porteus Maze test. Anxiety was considered both in terms of the relatively long-term, chronic anxiety represented by neurosis, and temporary or acute stress induced by test interference during the experimental procedure. It was assumed that the effects of both kinds of tension would be manifested in behaviors measured by the *Q* score procedure.

Though there is disagreement regarding the etiology of neurosis, there is consensus that the neurotic is anxious and fearful (Dollard & Miller, 1950, p. 222; Freud, 1933, pp. 114-115; Hoch, 1950, p. 106; Horney, 1937, p. 23; Maier, 1949, p. 1950; Masserman, 1943, pp. 67-68; Mowrer, 1950, p. 537; Sullivan, 1953, pp. 313-316). In the present study, a sample of adolescents who had been diagnosed as neurotic was compared to a group of presumably normal adolescents, and it was hypothesized that the *Q* score (the number of qualitative errors) of the neurotic adolescent is higher than the *Q* score of the normal adolescent. It is generally agreed that under interfering test conditions or conditions of stress, the neurotic reacts less adequately and with more anxiety to the test situation than the normal (Mandler & Sarason, 1952; Sarason, Mandler, & Craighill, 1952). In this study, test interference, introduced in the second presentation of the mazes to both groups, was comprised of instructions suggesting subject inadequacy and test procedures involving distraction, conflicting buzzer stimuli and time pressure, factors acknowledged to be stressful (Lazarus, Deese, & Osler, 1952). It was hypothesized that under interfering test conditions designed to increase tension or anxiety, the *Q* scores of both normal and neurotic adolescents are higher than under noninterfering test conditions. Furthermore, it was hypothesized that

under conditions of test interference designed to increase tension or anxiety, the increase in *Q* score for the neurotic adolescent is greater than the increase in *Q* score for the normal adolescent.

In summary, the following hypotheses were tested: (a) The *Q* score of the neurotic adolescent is higher than the *Q* score of the normal adolescent. (b) The *Q* scores of both normal and neurotic adolescents are higher under interfering than under noninterfering test conditions. (c) The difference in *Q* scores between interfering test conditions and noninterfering test conditions is greater for the neurotic adolescent than for the normal adolescent.

#### METHOD

The subjects in the study were 30 adolescent boys from the Ridgewood Public Schools, the normal group, and 30 adolescent boys from The Devereux Schools, the neurotic group. The procedure involved two presentations of the mazes to the normal adolescent and the neurotic adolescent, first, under noninterfering test conditions and, second, under interfering test conditions.

The normal subject was defined as an individual in "good standing" in the Ridgewood Public Schools: that is, a person who had not been referred to the school psychologist for behavior problems or academic difficulties. The individual's cumulative school record did not present adverse teacher comment, nor did it present extreme variability in academic performance from year to year.

The neurotic subject was defined as an individual who was not brain damaged, mentally retarded, psychotic, or delinquent. The individual's performance in school and behavior in society had made it necessary for him to be removed from school and society, and placed in an institution, The Devereux Schools, which provides special education, environmental therapy, and psychotherapy. Absence of delinquency was defined in terms of absence of acting out behavior, absence of a delinquency diagnosis by the examining psychologist and psychiatrist, and absence of a court record of delinquency.

The normal and neurotic groups were equated for age, intelligence (Porteus, 1950, pp. 151-160), and parent socioeconomic level (Edwards, 1938, pp. 3-6). The subjects who met the requirements for inclusion in the normal and the neurotic groups were asked to participate in a study to determine whether the mazes were sensitive to individual differences. All the subjects who were asked to participate in the study agreed to do so.

#### *Experimental Procedure and Instructions*

The procedure involved two presentations of the mazes to the normal adolescent and the neurotic



adolescent, first, under noninterfering test conditions and, second, under interfering test conditions.

The subjects were seen individually and introduced to the testing procedure with the following statement: "I want to find out how well you can find your way through these mazes. There are 10 mazes in all, each one somewhat more difficult than the one before. Here is the first maze." The standard Porteus test procedure was then followed, beginning with test instructions to the Year Five maze (Porteus, 1950, pp. 155-157). However, each maze was presented to the subject until it was completed successfully, even though the number of trials exceeded the limit set in the standard procedure. The time required for the completion of each maze was recorded. The first presentation of the mazes, as described, is the noninterference presentation.

At the conclusion of the testing, the examiner proceeded, ostensibly, to score the mazes, indicating to the subject in the process that he had not done as well as he might have done. The subject then was told by the examiner:

You did fairly well that time, but perhaps you can do better this time. This time the faster you are able to go through the maze, the higher your score will be.

In order to help you go at a faster rate, I will press the buzzer once, like this, whenever you come to a place in the maze where you might go wrong. Attention to the buzzer should help you to avoid errors.

Also, in order to keep you from losing time points, I will sound the buzzer twice, like this, if I find that you are taking too much time going through the maze. If more than two time warnings are necessary, your paper will be taken from you and you will receive a score of zero.

Good luck. If you try hard, you should be able to improve your score.

The buzzer pattern for each maze was double buzzer, single buzzer, single buzzer, double buzzer. The two single buzzer stimuli followed the first double buzzer stimulus, and were completed before presentation of the second double-buzzer stimulus. Timing for the buzzer presentation began with the presentation of the maze to the subject and proceeded according to the following schedule:

1. For mazes Year 5 through Year 8, the first double buzzer was sounded at 3 seconds, the second double buzzer at 6 seconds.
2. For mazes Year 9 through Year 10, the first double buzzer was sounded at 3 seconds, the second double buzzer at 10 seconds.
3. For mazes Year 2 through Adult I, the first double buzzer was sounded at 3 seconds, the second double buzzer at 15 seconds.

The mazes were rotated 180 degrees on the second presentation. This procedure was followed in order to reduce the facilitating effects of practice during the first presentation of the mazes. The time required

TABLE 1  
CHARACTERISTICS OF NORMAL AND  
NEUROTIC GROUPS

Characteristics	Normal	Neurotic
Mean age	16.42	16.06
Standard deviation of age	1.52	1.74
Mean intelligence quotient	99.60	96.73
Standard deviation of intelligence quotient	11.96	7.33
Socioeconomic level of parents		
Professional	8	8
Managerial	18	18
Skilled workers and foremen	3	2
Clerks and kindred workers	1	1
Semiskilled workers	0	1

for the completion of each maze was recorded. The second presentation of the mazes, as described, is the interference presentation.

The mazes were presented to each subject twice: first under noninterfering conditions, and then under interfering test conditions. The subject's performance under the two conditions was compared.

### Modifications in *Q* Score

Two modifications of the errors contributing to *Q* score were made for this study in order to control the number of mazes used by each subject and two modifications of *Q* score were made as a result of clinical observation of the maze performance of normal and neurotic adolescents.

The number of mazes used influences *Q* score since the greater the number of mazes used, the greater the probability of *Q* errors. In order to compare the *Q* performance of normal and neurotic adolescents without the effects of the variable, Number-of-Mazes-Used, the 10 completed mazes were scored in determining *Q* score rather than both completed and incompleting mazes.

The use of the 10 completed mazes for *Q* scoring made it necessary to exclude first third and last third errors from the *Q* score, since these errors only occur on incompleting mazes.

Previous clinical observation revealed that a breakdown of motor coordination, particularly under test interference or stress, was a characteristic of the maze performance of the anxious person. The breakdown of coordination was reflected in the gross error: the cut corner with white space showing, and the broken crossed line. The gross error was, therefore, included in the *Q* score used in this study.

In summary, the errors comprising *Q* score for this study were modified by eliminating the first third and last third error, and further modified by adding the gross error: cut corner with white space

TABLE 2  
ANALYSIS OF VARIANCE OF NORMAL AND NEUROTIC ADOLESCENT  
Q SCORE UNDER INTERFERENCE AND NONINTERFERENCE

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	F <sub>.99</sub>	F <sub>.999</sub>
Between Normal and Neurotic	250.56	1	250.56	79.98**		11.97
Between Interference and Noninterference	48.22	2	24.11	6.83*	4.98	7.76
Individuals within groups	549.41	58	9.47	2.68*	1.79	2.69
Interaction	204.61	58	3.53	6.78**		3.41

Note.—The Q score is the sum of weighted errors on the 10 completed mazes divided by 10.  
\* $p < .01$ .  
\*\* $p < .001$ .

showing (two points) and broken crossed line (four points). The Q score used in this study is the sum of the weighted error points on the 10 completed mazes divided by 10. Two independent judges correlated .98 for reliability, scoring 120 mazes representing each maze age level, each subject, and each test condition.

TABLE 3  
FREQUENCY DISTRIBUTION OF NORMAL AND NEUROTIC ADOLESCENT Q SCORES UNDER INTERFERENCE AND NONINTERFERENCE

Scores	Normal non-interference	Neurotic non-interference	Normal interference	Neurotic interference
0- .9	2			
1.0- 1.9	4		3	
2.0- 2.9	7	4	7	1
3.0- 3.9	10	2	6	3
4.0- 4.9	5	4	6	6
5.0- 5.9	0	7	3	3
6.0- 6.9	2	5	2	4
7.0- .79		5	1	4
8.0- 8.9		1	1	5
9.0- 9.9		0	1	0
10.0-10.9		0		0
11.0-11.9		1		1
12.0-12.9		1		1
13.0-13.9				0
14.0-14.9				1
22.0-22.9				1

Note.—The Q score is the sum of weighted errors on the 10 completed mazes divided by 10.

## RESULTS

The first hypothesis stated that the Q score of the neurotic adolescent is higher than the Q score of the normal adolescent. Table 2 presents the analysis of variance of Q scores by normal and neurotic adolescent subjects by normal and neurotic adolescent subjects under test conditions of interference and non-interference. As indicated in Table 2, the hypothesis was supported beyond the .001 level.

The second hypothesis stated that the Q scores of both normal and neurotic adolescents are higher under interfering test conditions than under noninterfering test conditions. As indicated in Table 2, this hypothesis was supported beyond the .01 level.

The third hypothesis stated that the difference in Q scores between interfering test conditions and noninterfering test conditions is greater for the neurotic adolescent than for the normal adolescent. As indicated by the significance of the interaction term of the analysis of variance summarized in Table 2, the hypothesis was supported beyond the .001 level.

An examination of the frequency distribution of Q scores by normal and neurotic subjects under test conditions of interference and noninterference presented in Table 3 revealed relatively small overlap between neurotics and normals under both noninterfering and inter-

fering test conditions. For example, under noninterfering test conditions, 23 normal adolescents obtained a  $Q$  score of 3.9 or less while only six neurotic adolescents did so. Under interfering test conditions, 22 normal adolescents obtained a score of 4.9 or less while 10 neurotic adolescents scored 4.9 or less. Thus, while some overlap existed,  $Q$  score clearly differentiated normal from neurotic maze performance under both noninterfering and interfering test conditions, the neurotic adolescent committing more  $Q$  errors than the normal adolescent. Additional findings of this study were:

1. The neurotic adolescent spent less time completing the mazes than the normal adolescent. Also,  $Q$  score tended to increase as time required for maze completion decreased; thus, it seems that the use of less time for maze completion by the neurotic adolescent contributed to the higher  $Q$  error of the neurotic adolescent.

2. The  $Q$  error performance of the neurotic adolescent was more variable than the  $Q$  performance of the normal adolescent under both interfering and noninterfering test conditions. Also, the  $Q$  error performance of the neurotic adolescent under test interference was more variable than the  $Q$  performance of the neurotic adolescent under noninterfering test conditions while homogeneity of variance existed for the normal adolescent under the two test conditions. (In order to make allowance for the heterogeneity of variance on the form of the  $F$  distribution [Lindquist, 1953, p. 83], the level of significance for this study was set at the .01 rather than the .05 level.)

3. The  $Q$  error categories which contributed most significantly to the greater number of neurotic errors, the greater number of errors under test interference, and the greater increase in neurotic errors under test interference were the cut corner error, the crossed line error, and the gross error. The greatest increase in the number of neurotic errors under test interference occurred in the gross error category.

4. The first third error, the last third error and, particularly, the perseverative error in-

creased under conditions of neuroticism and test interference, thereby contributing to the increase in the number of mazes used under conditions of neuroticism and test interference. The power of the perseverative error was illustrated by an increase of 7.1 times in the neurotic mean perseverative error under conditions of test interference.

## DISCUSSION

This research began with the clinical observation that the performance of anxious persons differed from that of nonanxious persons in certain aspects of behavior during administration of the Porteus Maze test. The apparent differences were usually not reflected in the quantitative score but, rather, in the kinds of performance characteristics described by Porteus in his development of the  $Q$  score. Thus, this study was designed to test the validity of the clinical observation that  $Q$  score provides a sensitive and reliable measure of anxiety. The results of the research clearly support the initial clinical observations. The results also are the first empirical demonstration of an existing relationship between anxiety and  $Q$  score.

Studies concerning muscle tension, motor coordination, and control functions provide a framework within which the increase in  $Q$  errors under conditions of neuroticism and test interference can be understood. There is considerable evidence that muscle tension is characteristically associated with anxiety (Jost, 1941; Jurko, Hill, & Jost, 1952; Malmo, Shagass, & Davis, 1950), and that this tension interferes with relatively fine motor coordination (Arnold, 1942; Malmo, 1950, pp. 169-180). This study found that the greatest increase in qualitative errors under conditions of neuroticism and test interference occurred in the cut corner, crossed line, and gross error categories, categories particularly susceptible to poor motor coordination. Several studies also suggest that control of psychomotor functions tends to deteriorate under conditions of neuroticism, anxiety, and stress (Jost, 1941; King, 1954, p. 144; Malmo & Shagass, 1949). Thus, past and present research tends to support an



interpretation of the results of this study (more  $Q$  errors under neuroticism and stress) in terms of increased tension, disrupted muscular coordination in relatively fine motor tasks, and decreased control of motor impulses as a function of stress and anxiety.

Previous research suggests that variability of performance (Eysenck, 1944; Jost, 1941; Jurko, Hill, & Jost, 1952; Malmö & Shagass, 1949) and perseverative responses (Symonds, 1958, p. 15; Maier, 1949) are characteristic of neurotic behavior and neurotic performance under stress. This experiment demonstrated that  $Q$  score, in addition to clearly differentiating normal adolescent from neurotic adolescent maze performance under both noninterfering and interfering test conditions, also revealed performance characteristic of neuroticism, stress, and anxiety; namely, variability and perseveration of responses. These findings seem to indicate that use of the  $Q$  error as defined in this study under both noninterfering and interfering test conditions provides a unique method for the study of performance under varying conditions of anxiety. Furthermore, the results of the study seem to indicate that the  $Q$  error is a sensitive and reliable measure of anxiety, and may be used as a measure of neuroticism and performance under stress.

#### SUMMARY

The results of this experiment are the first empirical demonstration of an existing relationship between anxiety and  $Q$  score. Specifically, this research tested the following hypotheses: (a) the  $Q$  score of neurotic adolescents is higher than the  $Q$  score of normal adolescents ( $p < .001$ ); (b) the  $Q$  scores of both normal and neurotic adolescents are higher under interfering than under noninterfering test conditions ( $p < .01$ ); and (c) the difference in  $Q$  scores between interfering and noninterfering test conditions is greater for neurotic adolescents than for normal adolescents ( $p < .001$ ).

For the experiment, 30 neurotic and 30 normal adolescents were tested, first under noninterfering test conditions and then under

interfering test conditions. The two groups were equated for age, intelligence, and parental socioeconomic level.

#### REFERENCES

- ARNOLD, M. B. A study of tension in relation to breakdown. *J. gen. Psychol.*, 1942, 28, 315-345.
- DOCTER, R. F., & WINDER, C. L. Delinquent vs. non-delinquent performance on the Porteus qualitative maze test. *J. consult. Psychol.*, 1954, 18, 71-73.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- EDWARDS, A. M. *A social-economic grouping of the gainful workers of the United States*. Washington, D. C.: Government Printing Office, 1938.
- EYSENCK, H. J. The effect of incentives on neurotics and the variability of neurotics as compared with normals. *Brit. J. med. Psychol.*, 1944, 20, 100-103.
- FREUD, S. *New introductory lectures on psychoanalysis*. New York: W. W. Norton, 1933.
- HOCH, P. H. Biosocial aspects of anxiety. In P. H. Hoch, & J. Zubin (Eds.), *Anxiety*. New York: Grune & Stratton, 1950. Pp. 105-116.
- HORNEY, KAREN. *The neurotic personality of our time*. New York: W. W. Norton, 1937.
- JOST, H. Some physiological changes during frustration. *Child Developm.*, 1941, 12, 9-15.
- JURKO, M., HILL, T. S., & JOST, H. Pathology of the energy system. *J. Psychol.*, 1952, 33, 183-198.
- KING, H. E. *Psychomotor aspects of mental disease*. Cambridge: Harvard Univer. Press, 1954.
- LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, 49, 293-317.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Cambridge: Riverside Press, 1953.
- MAIER, N. R. F. *Frustration*. New York: McGraw-Hill, 1949.
- MALMO, R. B. Experimental studies of mental patients under stress. In M. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950. Pp. 169-180.
- MALMO, R. B., & SHAGASS, C. Physiologic studies in reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, 11, 9-24.
- MALMO, R. B., SHAGASS, C., & DAVIS, J. F. A method for the investigation of somatic response mechanisms in psychoneurosis. *Science*, 1950, 112, 325-329.
- MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 166-174.
- MASSERMAN, J. H. *Behavior and neurosis*. Chicago: Univer. Chicago Press, 1943.
- MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.

- PORTEUS, S. D. *Qualitative performance in the maze test*. Vineland, New Jersey: Smith Printing House, 1942.
- PORTEUS, S. D. Q-scores, temperament, and delinquency. *J. soc. Psychol.*, 1945, 21, 81-103.
- PORTEUS, S. D. *The Porteus Maze Test and intelligence*. Palo Alto, California: Pacific Books, 1950.
- SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 561-565.
- SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: W. W. Norton, 1953.
- SYMONDS, P. M. What education has to learn from psychology: VI: Emotion and learning. *Teach. Coll. Rec.*, 1958, 60, 9-22.
- WRIGHT, CLARE. The qualitative performance of delinquent boys on the Porteus Maze Test. *J. consult. Psychol.*, 1944, 8, 24-26.

(Received April 3, 1963)

## TWO MEASURES OF ANXIETY: A VALIDATION<sup>1</sup>

WILLIAM D. WINTER, ANTONIO J. FERREIRA, AND ROBERT RANSOM

*San Jose State College*

19 undergraduates were subjected to 2 conditions of low anxiety, 2 of experimentally produced high anxiety, and 2 of examination anxiety. During each situation, the Ss were administered the Affect Adjective Check List (AACL) and the Palmar Sweat Index (PSI), and later the Taylor MA scale. The findings were: (a) both the AACL and PSI validly reflected anxiety; (b) both yielded stable scores characteristic of the S; (c) PSI did not correlate with AACL; (d) MA scale correlated with AACL, but not with PSI; (e) PSI and AACL responded differently to examination anxiety; (f) on the PSI, certain fingers seemed to sweat more than others. These findings indicate a lack of congruence between these verbal and physiological anxiety measures.

Anxiety has been a core concept of every dynamic theory of behavior, and the measurement of anxiety has become a major task of scientific psychology. There have been two popular approaches to measuring anxiety: one involves the assessment of an individual's physiological functioning, or changes therein under stress; another, his verbalized self-description in interviews or written questionnaires. A good deal of effort has gone into developing adequate measuring instruments in these areas and into analyzing the theoretical meaning of their complex interrelationships (Martin, 1961; Sarason, 1960). One simple technique used to obtain a physiological correlate of anxiety is the Palmar Sweat Index (PSI), which has given promising results in a number of studies (Mowrer, 1953). Among the verbalized self-description techniques, the Taylor (1953) Manifest Anxiety (MA) scale, the semantic differential, and adjective check lists have been very popular. Recently, Zuckerman (1960) developed an Affect Adjective Check List (AACL) that permits a quick self-description of degree of anxiety felt. He presented evidence of the validity of this technique, stating, for example, that the anxiety measured "today" on the AACL of a group of college students increased on the day of an examination, and that the increase was related to fear of failing the examination (Zuckerman, 1962).

The general purposes of the present study were to determine whether these two different measures of anxiety, the PSI and AACL, reflect experimentally induced anxiety, and to what extent they overlap in what they are measuring.

### METHOD

#### *Materials and Subjects*

The procedure used for obtaining the PSI was based on Mowrer (1953) and on our own investigations of this technique (Ferreira and Winter, 1963). Briefly, each S received a 3 × 5 inch piece of Whatman #1 filter paper impregnated with 5% tannic acid, and a bottle with a dropper containing a solution of ferric chloride in acetone. The preparation of the chemicals followed the Mowrer description. The Ss were instructed as a group in the use of these materials, with suitable demonstrations. They painted each of the four fingers of their own left hands with a drop of the ferric chloride solution and spread it over the finger tips with the aid of the dropper. They waited 30 seconds for the solution to dry and then placed their painted fingers on the filter paper for a period of two minutes, attempting to maintain a constant pressure. For each finger print, the area of maximum precipitation (caused by sweating) was measured by a specially built densitometer following Mowrer's design. The S's PSI score for each experimental condition was the sum of the densitometer readings for his four finger prints.

The AACL was administered and scored in the standard manner described by Zuckerman (1960), except that the S was asked to check the adjectives which described how he felt "right now." This alteration of the instructions seemed appropriate since we were primarily interested in the S's response to immediate, experimentally induced anxiety. Each S was also given the Taylor MA scale.

The Ss in this experiment were 19 undergraduates (almost all sophomores and juniors) enrolled in a

<sup>1</sup> This study was partially supported by National Institute of Mental Health grant MH 06560-01. The authors wish to thank Charles E. Smith for his invaluable technical assistance on this project.



course dealing with elementary research methods in psychology. The Ss appeared well motivated and cooperative. There were 13 boys and 6 girls, all but one in their early twenties.

### Design and Procedures

The Ss were exposed to six experimental conditions designed to differ in the amount and content of provoked anxiety. These were not meant to be equidistant along a scale of predetermined anxiety. The six conditions consisted of two conditions of low anxiety, two of examination anxiety, and two of experimentally provoked high anxiety. The six experimental conditions were given 3 to 14 days apart. Two weeks after the last session, the Ss took the Taylor *MA* scale. Except for an occasional absentee, all Ss were tested in a group at the same time of day by the same Es.

Briefly described, the six experimental conditions were:

I. *High Anxiety (Instructions)*. The Ss were instructed in the PSI technique in a very formal, unsmiling manner by an E who was a stranger to them. The real danger inherent in spilling the ferric chloride solutions on their clothing was stressed, and the PSI was described as a test to see how tense or nervous they were. This approach worked quite well; the Ss seemed very sober and worried while learning to manipulate the PSI materials and were concerned whether their own prints were too dark.

II. *Low Anxiety (Neutral)*. E was friendly, congratulated the class on handling the materials with no accidents in the previous session. The actual purpose of the study was explained and stress was placed on our interest in the effects of the experimental procedures on the PSIs of the group as a whole. E stated that the individual PSI of a S did not reflect his personality, and that we were repeating the PSI now to check on the reliability of the instrument and its correlation with AACL. The Ss appeared more relaxed and asked several intellectual questions about the project.

III. *Examination Day*. The PSI and AACL were administered just prior to the first regularly scheduled class examination.

IV. *High Anxiety (Lobotomy film)*. The E made several vivid remarks about the unpredictable results of lobotomy and the dangers of its unrestrained use. Then a film, "Prefrontal Lobotomy," was shown. At the point when the surgeon began to sever the neural fibers and the squirming and groans of our class seemed to reach a peak, the film was stopped and our measurements taken.

V. *Low Anxiety (Pleasure film)*. The Ss were told to relax, that they had a treat in store for them for being such good Ss. They were shown a humorous pantomime film, "The Chairy Tale," which involves a good deal of tension reduction at the end of the film, just before the PSI and AACL were administered. Again, audience response was as expected.

TABLE 1  
SUMS OF RANKS AND SIGNIFICANCE LEVELS  
FOR FRIEDMAN TESTS

	Anxiety conditions						$\chi^2$
	High		Exam		Low		
	I	IV	III	VI	II	V	
PSI	39.5	45	84	65	75	90.5	31.2*
AACL	81.5	38	56	43	93.5	87	41.7*

\*  $p < .001$ .

VI. *Examination Day*. The second regularly scheduled classroom examination.

For the two conditions involving the use of films, the fingers were painted before the films were begun and the prints taken when the film was stopped, thus avoiding a time lapse that might have permitted the emotional state to dissipate while the S was manipulating the PSI materials.

The basic hypothesis tested was that the S's scores on the PSI and AACL would vary systematically from condition to condition. We also tested (a) the correlations between PSI, AACL, and *MA* scale; (b) the consistency of the scores of a given S from condition to condition; (c) the relationship of our examination situations to the other anxiety conditions; (d) and whether the four fingers of our Ss were equally sweaty.

### RESULTS

The PSI and AACL scores of each S on the six experimental conditions were ranked in terms of the degree of anxiety shown, with the rank "one" representing the greatest anxiety. Friedman two-way analyses of variance (Siegel, 1956) were computed and the results were very significant (Table I), indicating that for both PSI and AACL, the Ss' responses were influenced by the differential effects of the six anxiety conditions.

The Ss were rank ordered within each condition according to their PSI and also their AACL scores. Kendall's *Ws* for the PSI rankings on the six conditions ( $W = .682$ ,  $p < .001$ ) and for the AACL rankings ( $W = .319$ ,  $p < .02$ ) indicated that each S tended to keep his same relative position in the group in all six conditions. In other words, both measures reliably differentiated the Ss, inasmuch as the changes in the S's scores tended to be of the same magnitude across

conditions, so that his ranking relative to the other *Ss* remained the same. Although both *Ws* were statistically significant, it should be noted that the PSI revealed a greater "concordance," or agreement from condition to condition, than did the AACL.

The two measures did not correlate significantly with each other within any single condition or across all conditions ( $r_s = +.03$ ). One factor involved in this lack of correlation can be discerned by referring to Figure 1, which presents the median values on each of our tests for the combined conditions of high (I plus IV), examination (III plus VI), and low (II plus V), anxiety. It will be seen from this figure that on the AACL the *Ss* treated the examination days as similar to the high anxiety conditions, and both of these as significantly different from the low anxiety conditions. However, the examination day scores on the PSI approximated those of the low anxiety conditions, and both differed significantly from the high anxiety situations. All four of these significant differences were at the  $p < .001$  level by the sign test, two-tailed. Apparently, for our sample, the verbal and physiological measures of anxiety responded differently to the tension produced by classroom examinations.

We also found that the combined AACL scores of our *Ss* for all six conditions correlated significantly with their scores on the Taylor *MA* scale ( $r_s = +.44$ ,  $p < .05$ , one-tailed). This had been predicted on the basis of Zuckerman's findings (Zuckerman, 1962).

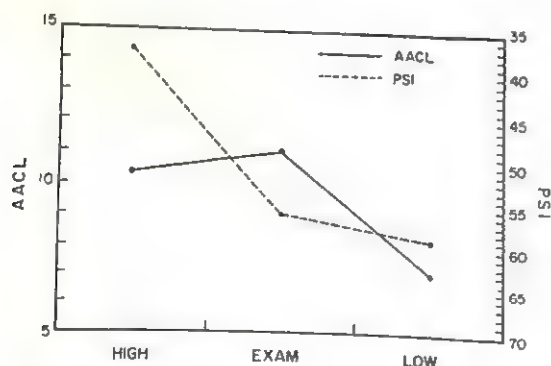


FIG. 1. MEDIAN VALUES OF PSI AND AACL FOR THE HIGH, EXAMINATION, AND LOW ANXIETY CONDITIONS. (THE DIRECTION OF PSI SCORES IS REVERSED SO THAT BOTH ORDINATES MEASURE INCREASING ANXIETY.)

There was no correlation between the PSI and *MA* scale. No significant sex difference was found on either measure, but the small number of female *Ss* in the sample precludes adequate statistical analysis of this variable.

Two other interesting findings on the PSI should be mentioned. In a previous study investigating the methodological problems of the PSI technique (Ferreira & Winter, 1963), we found that the *Ss'* fingers did not sweat equally but that some fingers produced darker prints than others. To test the generality of this finding with our present sample, the PSIs for the four fingers of every *S* were rank ordered in all six conditions and a *W* computed for each *S*. The sum of the chi square values for all *Ss* combined was very significant ( $\chi^2 = 116.89$ ,  $df = 57$ ,  $p < .001$ ), indicating that, as in the previous study, the *Ss'* sweat production varied from finger to finger in a way perhaps consistent for, and characteristic of, the given individual.

The sums of the PSI scores obtained for each finger on all conditions combined were ranked, and a *W* computed for all *Ss*. The resulting rank ordering was significantly different from chance ( $W = .19$ ,  $p < .02$ ), showing that, for our sample, the third or ring finger of the left hand was the most sweaty and the fourth and index fingers the least sweaty (only left hands tested).

## DISCUSSION

It seems from our results that both the PSI and the AACL reflect the different experimental conditions of anxiety, and that both can be considered simple, practical, and valid measures of anxiety. In addition, both yielded stable scores characteristic of the *S*. However, the lack of correlation between the PSI and AACL, their different response to examination anxiety, and their different correlations with the Taylor *MA* scale suggest that these two instruments may be measuring different phenomena (Rosenstein, 1960; Sarason, 1960). This interpretation is congruent with the results of McGuigan, Calvin, and Richardson (1959), who also found no correlation between the *MA* scale and the PSI.

We are still faced with the important



question of just what these various instruments actually do measure and how they relate to one another. We have some speculations in the literature, but little definitive research. One theoretical possibility, of course, is that anxiety is not a unitary concept. Eysenck (1961), later supported by Buss (1962), has pointed out two factors in anxiety: one, a somatic factor of autonomic reactivity, and the other, a psychic or ideational factor. The PSI probably represents one measure of Eysenck's somatic factor. The AACL score probably measures the psychic factor of anxiety, but not directly, since it also reflects whether the *S* is willing and able to communicate the ideational content of his anxiety in an undistorted fashion. Even if the PSI represented pure somatic anxiety and the AACL pure psychic anxiety, we would generally expect them to covary (as in our *high* versus *low* anxiety experimental situations), but on some occasions to be uncorrelated or even negatively correlated (e.g., our examination days).

As stated above, part of the discrepancy between PSI and AACL is probably due to the communication variable on the AACL, i.e., to the *S*'s ability and willingness to admit his anxiety to himself or others. For example, to the *S* it might be socially acceptable, or even expected, to admit publicly on the AACL that he is anxious on an examination day. But in Condition I, when each *S* may be privately worried about what his PSI might indicate about his personality, the typical AACL response may be a bland, defensive denial of anxiety, even if his PSI is elevated. Although there appears to be some truth to this approach, it does not explain why the PSI showed little anxiety on examination days. This particular finding would seem to contradict the results of other investigators (Beam, 1955; Davis, 1957) who found that the PSI increased just before an examination. This inconsistency may be due to situational variables such as how the examinations were viewed, the severity of test anxiety, etc. Confirmation of this possibility cannot come from our data, but it is of interest that when our *Ss* were shown these group results after the study was completed

and asked to discuss them, many stated that they had not been really very worried about the examinations which came so early in the course with an "easy-going instructor" (*sic*).

Some writers like McGuigan, et al. (1959), have suggested a somewhat different explanation for the lack of correlation between verbal and physiological anxiety measures, specifically the PSI and *MA* scale. They state that the PSI is a better measure of situational, temporary anxiety, while a questionnaire like the *MA* scale is a better measure of anxiety as a stable characteristic of personality. Presumably, the *MA* scale reflects the verbalized self-concept, which may be more permanent than the physiological imbalance reflected in the PSI. Again, there is probably some truth in this idea, but it should be remembered that in our study the PSI scores for a given *S* were generally stable and characteristic of him in all conditions. Also, our AACL instructions stressed how the *S* felt right at the moment of testing, rather than how he generally feels, which is the usual set for the *MA* scale. Therefore, the stability of our PSIs and the "immediacy" of our AACL scores would seem to make the explanation of McGuigan et al., inappropriate in this instance. In fact, we can only conclude our discussion of the lack of correlation between the AACL and PSI by stating that no simple, adequate explanation presently exists for the complex interrelationship of these two anxiety measures.

The remaining finding, that for every *S* some fingers are characteristically sweatier than others, agrees with our previous results with a different sample (Ferreira & Winter, 1963). This has not been noted previously in the PSI literature, and it opens some important questions that may well justify the efforts of future investigators.

#### REFERENCES

- BEAM, J. C. Serial learning and conditioning under real life stress. *J. abnorm. soc. Psychol.*, 1955, 51, 543-551.
- BUSS, A. H. Two anxiety factors in psychiatric patients. *J. abnorm. soc. Psychol.*, 1962, 65, 426-427.
- DAVIS, R. H. A further study of the effect of stress on palmar prints. *J. abnorm. soc. Psychol.*, 1957, 55, 132.



- EYSENCK, H. J. *The handbook of abnormal psychology*. New York: Basic Books, 1961.
- FERREIRA, A. J., & WINTER, W. D. The palmar sweat print: A methodological study. *Psychosom. Med.*, 1963, 25, 377-384.
- MARTIN, B. The assessment of anxiety by physiological behavioral measures. *Psychol. Bull.*, 1961, 58, 234-255.
- MCGUIGAN, F., CALVIN, A., & RICHARDSON, ELIZABETH. Manifest anxiety, palmar perspiration-index, and stylus maze-learning. *Amer. J. Psychol.*, 1959, 72, 434-438.
- MOWER, O. H. *Psychotherapy: Theory and research*. New York: Ronald Press, 1953.
- ROSENSTEIN, A. Psychometric versus physiological anxiety and serial learning. *J. Pers.*, 1960, 28, 279-292.
- SARASON, I. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, 57, 403-415.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- TAYLOR, JANET. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- ZUCKERMAN, M. The development of an Affect Adjective Check List for the measurement of anxiety. *J. consult. Psychol.*, 1960, 24, 457-462.
- ZUCKERMAN, M. Further data relevant to the construct validity of the Affect Adjective Check List measure of anxiety. *J. consult. Psychol.*, 1962, 26, 291.

(Received April 9, 1963)

## THE RORSCHACH BARRIER SCORE AND SOCIAL BEHAVIOR<sup>1</sup>

JOHN RAMER<sup>2</sup>

*University of Washington*

96 undergraduate females placed in 3 experimental situations were assessed with regard to quantity and quality of communication each sent. The situations were those of no feedback, negative feedback, and positive feedback. Their communication behavior was related to a Rorschach content score, the barrier score. It was found that high barrier scores would attempt more communication, assert themselves more, tend to be less self-deprecating, and tend to express less discomfort in a situation increased in degree of threatening quality.

Although many studies have attempted to relate Rorschach scores to other test scores and other kinds of behavior, none has enjoyed the success of the barrier score. The relationships obtained have been many and various (Cleveland, 1957, 1960; Davis, 1960; Fisher & Cleveland, 1956a, 1956b, 1957, 1958a, 1958b; Fisher & Fisher, 1959). This score, the barrier percept, refers to all percepts on the Rorschach test having to do with the assignment of definite structure, definite substance, and definite surface qualities to the bounding peripheries. The authors of the barrier score, Fisher and Cleveland, conceived of it as a projection of the person's bodily organization and that the individual perceives his body exterior as having (or lacking) protective characteristics. Controlled experimental studies relating barrier score and social behavior have not been available.

Viewing the barrier score as an unconscious projection of a person's definitive existence and his felt ability to adequately cope with the demands of reality, a communication task was created to test three specific hypotheses. They were: (a) that high barrier persons would show a greater frequency of communication attempts in an interpersonal situation than the low barrier persons; (b) that

a content analysis of the communication occurring in an interpersonal situation would show high barrier persons manifesting more committal and directive statements, more disagreement, less passive-acceptance statements and fewer self-deprecating statements; and (c) that qualitative differences would obtain greater magnitudes in a frustrating and/or threatening situation as opposed to a neutral situation.

### METHOD

#### *Subjects*

The final subject population for this study was 96 females acquired from Seattle University, a co-educational Catholic school, and the University of Washington, a large state school. Forty-eight subjects were drawn from each school and all except five of the total subjects were gotten from women's living groups.<sup>3</sup> The one group was acquired on a completely random basis with the help of the Dean of Women. The other group was gotten on a completely volunteer sign up basis using sheets posted on the dorm bulletin boards. Only freshmen and sophomore girls were used.

#### *Task and Apparatus*

The communication task demanded discussion of written stories with an assumed partner via slips of paper that the experimenter was supposed to deliver. For this task four subjects at one time were individually seated at card tables, the four of which were arranged as along sides of a rectangle, with the experimenter in the middle. The upper halves of their bodies were surrounded by three-sided screens so as to completely block each subject's view of the other three subjects. Before each of them was a set of three pictures, two pencils and a large stack

<sup>1</sup> This article is based on a dissertation submitted to the faculty of the Department of Psychology of the University of Washington in partial fulfillment of the requirement for the PhD degree. The author wishes to acknowledge his indebtedness to Ezra Stotland, Seymour Fisher, and A. D. Davis for their guidance and support.

<sup>2</sup> Now with the Department of Psychology, Western State Hospital, Fort Steilacoom, Washington.

<sup>3</sup> Five subjects were solicited from a psychology of adjustment class.

of blank slips of paper. The pictures were identical to the ones used by Kurt Back in his classic study (Back, 1951).

Additional material consisted of standard prepared messages. These messages were basic to the controlled experimental situations. They were varied as to their threatening quality and each group of four subjects found themselves in one of three situations.<sup>4</sup> They might have received no replies. They might have received positive-neutral replies, e.g., "I'd really like to know about your story," "I'm sorry I'm not any more help than I am, but I really appreciate your messages," and "I have a feeling my story agrees with yours, but I'm sure you're a much better writer than I am." Or thirdly, they might have received negative replies, e.g., "I hope we will have the same ideas for our stories," "I have a feeling our stories are probably different," and "Since I've had some experience writing this kind of story before, I'm really satisfied with what I've written." All subjects received the same number of messages and at a similar time.

### *Procedure and Instructions*

Subjects were run in groups of four. Each subject upon arrival at the experimental room was seated at a card table behind a screen and was told the experiment would begin in a few minutes as soon as all four subjects were assembled. The groups of four were successfully chosen so as to eliminate complete familiarity with one another. When all were seated these directions were read:

You are being asked to partake collectively in writing a group story, each group to consist of two persons. For this you have been given identical sets of three cards. This study is part of an effort to predict which groups of two, out of the 48 groups I will eventually test, can accomplish the best results—please do your best.

First, each of you has been given the same set of pictures to be used in the final study and each of you will be allowed 7 minutes to write a short but comprehensive story with these directions. Your task is to write a story from the set of three photographs which depict quite a commonplace incident. This gives you an opportunity to give play to your imagination, although the story should be plausible and supported by features of the pictures. The pictures themselves, being taken from a film strip, form a sequence which you will have to reconstruct. Then you will write a story connecting the pictures. Remember—you should write a good story but it is important to make it plausible by use of the available clues. You should pay particular attention to the setting of the story, the relationship of the two people involved and the plot of the

story. These criteria will be used for the evaluation of the stories. You may begin!

The subjects were told when 2 minutes remained and when 1 minute remained so they could put endings on their stories. When the time had expired the experimenter collected both the written stories and the sets of pictures. Then he introduced the next part with these instructions.

The next part will be a group discussion with your partner. Since I could not hope to write and record as fast as you all talk, especially with two groups at a time, this will take the form of a paper and pencil discussion. At my signal you may initiate your story discussion with your partner, whom I have randomly chosen, by sliding messages under your front board. You will notice a groove in the middle of the board. I will then carry your message to your partner. The slips of paper in front of you are to be used for writing your messages, however, each slip of paper should contain only one sentence or complete thought and please put your identity number in the upper right corner. Of course you may talk about any and all aspects of your story or simply make talk. The partitions are used to conceal the identity of your partner and to prevent you from observing any facial grimaces, smiles, and other behavioral signs. Remember—how many notes you send is entirely up to you. Some may wish to send just one note, some a few notes and some may send no notes at all. If you receive a note from someone and don't wish to respond to it, you may answer with whatever you wish. Also, messages may be in transit and not apparently answered, or you may need to send several messages before your partner sees your point clearly. If there are no further questions—you will now have seven minutes for your discussion with your partner. You may begin!

When this was over the subjects were asked to write their stories a second time without the aid of the pictures and told they could make use of their group discussions. This additional data was not pertinent to the present study, however, and is not presented. A short postsession questionnaire was then filled out by each subject. It attempted to ascertain the efficacy of the experimental technique. The subjects were then told what had transpired and asked not to discuss the experiment until further word from the experimenter.

### *Barrier Score*

Each subject was seen a second time approximately 2 to 3 weeks later on a chance basis with reference to groups and situations, and administered a group Rorschach. The number of responses were controlled. There was no formal inquiry but only these directions.

This can be considered an exercise in perception. You are going to be shown a series of slides

<sup>4</sup> These three situations, the zero replies, the positive replies, and the negative replies, are hereafter referred to as the Z-Situation, the P-Situation, and the N-Situation, respectively.



each of which contains the picture of an inkblot. This is a modification of the inkblot test or the Rorschach as it is sometimes called. I would like you to look at each inkblot and write down the first three things it looks like or reminds you of, or makes you think of. For some inkblots I will ask for just the first two things you are reminded of. I am asking for just a set number of answers because I know from experience that people can see these many things. However, people see a great many things in these inkblots and so there are no correct answers. Please describe as adequately or as fully as you can your perceptions and/or what about the blots reminds you or make you think of what they do. If I go too fast or you have any questions, please raise your hand. Also let me know if it is impossible for you to see the screen.

The response total was limited to 25 responses based upon the card pull values as indicated by Grayson's study (Grayson, 1956). The first three responses were asked for Cards I, II, III, VIII, and X. The first two responses were asked for the remaining cards. The experimenter introduced each card saying, "this is Card number x and for this card I would like just the first three (or two) things this inkblot reminds you of or makes you think of." Three minutes seemed sufficient for the two-response cards while 5 minutes were sometimes needed for the three-response cards. All protocols were typed with carbon copies allowing three independent judges to determine barrier scores.<sup>5</sup> The median barrier score was determined for each experimental situation and those above the median were designated as high barrier persons while the others were designated as low barrier persons.

## RESULTS

### *Barrier Score*

As can be seen in Table 1, the interjudge scoring of the barrier score was quite adequate with the average Pearsonian correlation being  $r = .89$ . This lends further support to the belief that the barrier score can be reliably assessed. The distribution of the barrier scores ranges from 0 to 15. The mean for the total distribution was 6.31, with the median being 5.82. An analysis of variance of the barrier scores for the three experimental conditions revealed no significant results.

Of greater importance was the analysis of the two university samples of subjects with regard to their barrier score distributions.

<sup>5</sup> Those scores determined by Seymour Fisher who was one of the three judges were used as criterion measures. The evaluation of responses as barrier scores followed Fisher's standards.

TABLE 1  
INTERJUDGE RELIABILITY OF THE  
BARRIER SCORE<sup>a</sup>

	J1	J2	J3
J1	—		
J2	.96	—	
J3	.81	.90	—

<sup>a</sup> The scores were determined for 96 protocols and the correlations are product moment.

A *t* test analysis resulted in no statistically significant difference. These results allow one to consider both samples as being selected from the same general population.

### *Assessment of Technique*

The success of this study was very dependent on the experimental subjects perceiving their fictional partners as real. Three questions in the postquestionnaire are relevant.

Question two asked, "How well do you know the other members—much, moderate, slight, not at all?" It revealed the fact that NO subject admitted to knowing the other members very well. For the most part they either did not know the other members or had only seen them around. They may have known one of the other members or even two of the other members to some degree but in no case did they know all the other members. The desired effects were apparently accomplished, meaning that the subject could not logically rule out all members of her group as the possible communicant.

Question three, "Did you know who your partner was and how well do you know her?," yielded different results according to the experimental conditions. Among the total 64 subjects in the N-Situation and P-Situation no one knew who their partner was although four subjects in the P-Situation chose an actual person in her group. In the Z-Situation there were 8 subjects out of the total 32 that expressed doubts about having had a partner. They wondered if the experimenter was delivering the messages. This is somewhat logical behavior since they received no replies at all. Interestingly enough, 6 of the 8 subjects who expressed this awareness were high barrier persons.

Question six asked the subjects what they thought the experiment was about and was another effort to uncover an awareness by the subjects of their fictive partners. Needless to say, no one knew what the experiment was about and only those who had previously expressed doubts about their partner did so again.

Of equal importance was the affective values of the standard preprepared messages. Question five, "Would like to meet and know your partner—why?," was an attempt to evaluate the effects of the three experimental conditions and the messages in particular. In the N-Situation there were only four subjects who did not care to meet their partner and learn to know her better. Unfortunately, the subjects did not verbalize exactly why they felt the way they did. The four answers were: "not necessarily so," "frankly I don't care," "I thought her notes senseless," and "we have different interests." All four of these subjects who verbalized this ability to reject their partners were high barrier persons. In the P-Situation there were two high barrier persons and one low barrier person who expressed a disinterest in meeting their partner. The high barrier answers were to the effect that "maybe, she doesn't seem to have great confidence," and "not more than the others for she didn't seem interesting." The low barrier person simply said, not especially. In the Z-Situation six low barrier persons expressed a disinterest in meeting and getting to know their partner better, while only one high barrier person did not wish to meet her partner because she thought she already knew her. This completely unstructured situation seemed to be much more upsetting to the low barrier person. Some of the answers were: "makes no difference," "not particularly, since she was not very sociable," and "doesn't matter," (this person said she felt frustrated by the whole situation). All of the other subjects in the three situations answered that they would like to meet their partner, with the usual answers being that they liked to meet people or to see what others wrote for their stories. The subjects in the P-situation were the only ones to describe their partner to any degree. The simulated partners were

described as friendly, vague, evasive, and some thought the partners lacked confidence.

### *Analysis of Hypotheses*

Hypothesis I predicted that the high barrier persons would attempt more communication than the low barrier subjects. Two measures of attempts to communicate were used.

The first measure used was simply the number of notes sent by the subjects to their simulated partners. An analysis of variance of these data is presented in Table 2. As can be noted there is a very significant difference between high and low barrier groups. The mean frequency of messages sent by the low barrier group was 3.77, while the high barrier group sent a mean of 5.12 messages. There was no significant interaction between barrier groups and situations. However, Table 2 does reveal that the situations exerted an influence on the frequency of messages. The average frequency sent in the Z-Situation was 3.84, in the N-Situation it was 5.12, and in the P-Situation the mean frequency was 4.69. The *t* tests indicated a significant difference only between the Z-Situation and the N-Situation. One could expect that no answering replies would tend to depress the amount of intercommunication.

Although the subjects were directed to write only one sentence to a slip of paper, this was not the result. Since slips of paper frequently contained several points for discussion, another measure of attempted communication was used. This consisted of the number of units of communication scored

TABLE 2  
A COMPLETE ANALYSIS OF VARIANCE FOR A  
3 × 2 FACTORIAL DESIGN

Source	df	F of units		F of messages	
		MS	F	MS	F
Between cells	5	25.11	1.96	11.94	2.76*
Situations (A)	2	34.64	2.70	13.58	3.14*
Barrier scores (B)	1	52.51	4.10*	31.51	7.28*
B × A	2	1.88	0.15	0.51	0.12
Within cells	90	12.81		4.33	
Total	95				

\*  $p \leq .05$ .

\*\*  $p \leq .01$ .



under the Bales system (Bales, 1950). In other words each subject's score was the number of units of communication instead of the frequency of messages. Admittedly, the number of units of communication is dependent on the number of slips of paper sent and a Pearsonian correlation of  $r = .79$  was obtained for the two measures. Table 2 also shows the results of an analysis of variance for the frequency of units. Once more confirmation of Hypothesis I is found. High barrier subjects sent a mean of 8.12 units, while low barrier subjects sent a mean frequency of 6.64 units. This represents a significant difference at the .05 level. There were no significant differences among the situations nor were there any significant interactions.

Hypothesis II predicted that high barrier subjects would give more directions and suggestions, make more committal and orienting statements, be less self-depreciating, and be able to disagree more frequently than the low barrier subjects. In terms of the Bales' categories for content scoring, Hypothesis II would be that high barrier persons would show more of Category 4 (gives directions), would show more preponderance of Category 5 and 6 (gives opinion and orientation) over Categories 7 and 8 (asks for orientation and opinion), would show less of Category 11 (self-depreciatory) and show less of Categories 1 and 3 combined ("going along" and passive acceptance).<sup>a</sup>

In order to test this hypothesis, chi square analyses were computed for the presence or absence of the selected categories between the communication efforts of the high and low barrier persons. The results revealed just one of the chi squares reached a statistically significant level and it indicated that the high barrier subjects as a group did more giving opinion and orientation as opposed to asking for opinion and orientation than did the low barrier group. However, all of the chi squares are in the predicted direction. High barrier subjects did tend to show more of category four and less of the behavior classified in

category eleven and in the combined category one and three.

Hypothesis III predicted that a frustrating and/or threatening situation would tend to increase the differences predicted in Hypothesis II. In terms of this study this meant that the differential frequencies for predicted Bales' categories would be heightened in the N-Situation and the Z-Situation. A chi square analysis broken down for the three experimental situations revealed only one chi square that reached a significant statistical level and it was in the N-Situation. However, the chi squares were greater in magnitude in the N-Situation for three of the four criterion categories and Z-Situation yielded the largest chi square for the lone category not accounted for by the N-Situation. It is difficult, just the same, to cite the present data as evidence for increased differences as due to the different situations. The frequencies within the chi square cells are so small and show such slight differences that one could easily question the stability of the present results. This is particularly true for the Z-Situation. One might describe the evidence as slightly suggestive.

#### DISCUSSION

In general, the hypotheses in this study found some confirmation. In a controlled communication situation the high barrier persons did show significantly different social behavior from the low barrier persons. There appeared to be both a quantitative and qualitative difference. The inability to demonstrate more than mere trends for some of the predictions can be seen as a partial result of the type of analysis undertaken. Originally, it was planned that the data would be analyzed by an analysis of variance technique. However, due to the lack of interjudge reliability for the individual units using the Bales system, this technique had to be discarded. The plan then involved considering the already scored communication data and deciding on the presence or absence of each category. This approach did yield reliable judgments; however, scoring for the absence or presence of a category naturally neglected the amount of presence. It can be seen that this type of analysis would weaken the

<sup>a</sup> Since all subjects showed the presence of categories 5, 6, 7, & 8, combinations were formed such that the prediction was that high barrier subjects would show categories 5 & 6 > 7 & 8, to a greater extent as a group than the low barrier subjects.



chances for any kind of significant results although significant results were noted.

Another weakness in this study was its failure to clearly demonstrate effects for the three experimental situations. Several pilot studies were done in an effort to construct positive and negative preprepared messages. Since no one was forced into the experiment it was hard for the students to accept some of the experimenter's more hostile messages. The subjects did not believe any female nice enough to take part would say such things. It also became apparent that all feedback would have to be rather vague and general so as to be universally applicable. In terms of real constructive help both sets of messages were negative. The final criteria for the messages were that they were successful in pilot studies to the extent of being plausible replies for the messages sent and elicited some of the desired effects although not totally so. The subjects were very hesitant in directing emotional replies to their fictional partner. In the postexperimental session feelings were verbalized but this was not true for the paper and pencil discussion. This was particularly evident in the pilot studies. Possibly if the subjects had not expected to ever meet their partners they would not have been so constricted. Or, if the subjects would have been paid a goodly amount of money they might have been more willing to express some feelings toward their partner who was now being paid to produce and was not. A monetary incentive might have helped greatly.

It would seem that body image research and the barrier score in particular demand further investigation. The question of what does the barrier score really measure is still in the process of being answered. There are some who feel the barrier score is another cognitive style much in the Kleinian sense. The barrier scorer in a way similar to the "leveller-sharpener" (Klein, 1959) supposedly is oriented to distinguishing boundaries of objects and/or situations. The meager evidence bearing on this problem does not support such an interpretation. However, this possibility has not been sufficiently investigated. Holtzman (1961), on the basis of his series of factor studies, noted that the barrier score

consistently clustered with indexes of the construct of "ego integration." If this construct manifests itself by an active willingness to cope with the demands of a situation in a purposive and meaningful way even under somewhat trying circumstances, then this study can also be seen as lending support to such an interpretation for the barrier score.

Still there is no denying that the barrier score significantly relates to unique physiological processes consonant with a body image theory. More recent studies by Fisher and Fisher have revealed a strong relationship between the barrier score and concern for body areas on an internal-periphery basis as indicated by introspective report. Judging from the literature about the relationship of body-images and personality integration (Arieti, 1955, Cornelison and Arsenian, 1960, Secheyay, 1951) the need is certainly recognizable for valid quantification in this area. If the barrier score does measure an aspect of a person's self-image, one should be able to predict more meaningfully what a high or low scoring person's behavior would be like in certain situations. Realistically speaking, within the area of body-image research the surface or barrier is just being scratched, or should one say penetrated.

#### REFERENCES

- ARIETI, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
- BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9-23.
- BALES, R. F. *Interaction process analysis*. Cambridge: Addison-Wesley, 1950.
- CLEVELAND, S. Prediction of small group behavior from a body image schema. *Hum. Relat.*, 1957, 10, 223-233.
- CLEVELAND, S. Body image changes associated with personality reorganization. *J. consult. Psychol.*, 1960, 24, 256-262.
- CORNELISON, F. S., & ARSENIAN, J. A study of the response of psychotic patients to photographic self-image experience. *Psychiat. Quart.*, 1960, 34, 1-8.
- DAVIS, A. D. Some physiological correlates of Rorschach body image productions. *J. abnorm. soc. Psychol.*, 1960, 60, 432-436.
- FISHER, S., & CLEVELAND, S. Relationship of body image boundaries to memory for completed and incomplete tasks. *J. Psychol.*, 1956, 42, 35-41. (a)
- FISHER, S., & CLEVELAND, S. Relationship of body image to site of cancer. *Psychosom. Med.*, 1956, 18, 304-309. (b)

- FISHER, S., & CLEVELAND, S. Body image boundaries and adjustment to poliomyelitis. *J. abnorm. soc. Psychol.*, 1957, 55, 88-93.
- FISHER, S., & CLEVELAND, S. Body image boundaries and sexual behavior. *J. Psychol.*, 1958, 45, 207-211. (a)
- FISHER, S., & CLEVELAND, S. *Body image and personality*. New York: D. Van Nostrand Co., Inc., 1958. (b)
- FISHER, S., & FISHER, RHODA. A developmental analysis of some body image and body reactivity dimensions. *Child Developm.*, 1959, 30, 389-402.
- GRAYSON, H. M. Rorschach productivity and card preference as influenced by experimental variation of color and shading. *J. proj. Tech.*, 1956, 20, 288-296.
- HOLTZMAN, W., THORPE, J., SWARTZ, J., & HERRON, W. *Inkblot perception and personality*. Austin: Univer. Texas Press, 1961.
- KLEIN, G., HOLZMAN, P., SPENCE, D., & LINTON, HARRIET. Cognitive control. *Psychol. Issues*, 1959, 1.
- SCHILDER, P. *The image and appearance of the human body*. London: Kegan, Paul, Trench, Trubner & Co., 1935.
- SECHEHAYE, M. A. *Autobiography of a schizophrenic girl*. New York: Grune, 1951.
- THIBAUT, J., & REICKEN, H. Determinants of social causality. *J. Pers.*, 1955, 24, 113-133.

(Received April 24, 1963)

## NOTES AND COMMENTS

### THE VALIDITY OF THE BARRON EGO STRENGTH SCALE IN AN INDIVIDUAL FORM

JEROME SILVERMAN<sup>1</sup>

*School of Medicine, University of Southern California*

This study investigated whether the Barron Ego Strength Scale by itself could demonstrate similar reliability and validity as when administered in the context of the full MMPI. 36 hospitalized normals and 32 chronic schizophrenics were administered the full MMPI and 1 week later an individual form of the *Es* scale. There was a very high correlation between *Es* scores on test-retest within each group. The construct validity of the *Es* scale was supported on both forms. It was concluded that the *Es* scale can be administered individually since removal from context of the full MMPI does not produce any significant changes in its reliability or validity.

Barron (1953) developed an Ego Strength scale (*Es*) which was validated against the predictive criterion of successful outcome in psychotherapy. The availability of a short, reasonably objective measure of ego strength can be a valuable tool for clinical and research purposes. Since the 68-item *Es* scale is extracted after administration of the entire MMPI, a problem arises for those interested in using the *Es* scale by itself. Different response sets may be elicited if items or single scales are isolated from a main body of items, thereby affecting the validity of the scale. This study investigated whether the *Es* scale by itself would measure the same thing as when the entire MMPI is administered.

#### METHOD

An individual *Es* scale was constructed comprising Barron's 68 *Es* items in addition to the 15 Lie items of the MMPI. Forty Veterans Administration male chronic schizophrenics were selected by the ward physician on the basis of ability to cooperate. Patients with signs of neurological involvement were excluded from the experiment. Another group consisted of 40 Veterans Administration male, non-psychiatric tuberculosis patients. These patients were selected by the ward physician as being "healthy psychologically." Patients with a history of psychiatric disorder; patients with physical complaints frequently termed "psychosomatic"—stomach ulcers, low back pain, asthma—were excluded from the experiment.

Both groups were administered the full MMPI and

one week later they were retested with the individual form. Subjects who obtained Lie scores of 8 or more on either the full MMPI or the Individual scale were not included in the statistical analysis. Four normals and eight schizophrenics were excluded from the sample due to their Lie scores.

The *Es* scores were then derived from both the Individual scale and the full MMPI.

The results as shown in Table 1 indicate rather clearly that scores obtained from the Individual *Es* scale closely approximate those *Es* scores obtained from the full MMPI. The product-moment correlations for the two forms were .91 for the normals and .85 for the schizophrenics. The mean *Es* scores of the two forms differed by only .2 for the normals and only .9 in the schizophrenic group.

If the *Es* scale were valid, one would expect the normals to have higher ego strength scores than the chronic schizophrenics. Table 1 indicates that the normals did show significantly higher *Es* scores than the schizophrenics on both

TABLE 1  
COMPARISON OF *ES* SCORES BETWEEN HOSPITALIZED  
NORMALS AND SCHIZOPHRENICS AS DERIVED FROM  
INDIVIDUAL FORM AND FULL MMPI

Es score from	Normal TB patients		Chronic schizophrenics	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
MMPI	43.50	7.31	37.30	6.30
Individual scale	43.70	7.84	38.20	7.86
	<i>r</i> = .91		<i>r</i> = .85	

<sup>1</sup> This study was carried out while the author was a Veterans Administration Clinical Psychology Trainee and a graduate student at the University of Pittsburgh.



the MMPI ( $t = 2.97 < .01$ ) and Individual forms ( $t = 2.43 < .02$ ). These results support the construct validity of the *Es* scale whether it be extracted from the full MMPI or Individual scale.

## REFERENCE

- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.

(Received April 17, 1963)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 6, 533-535

## EFFECT OF ORDER OF PRESENTATION OF TAT CARDS

LEWIS BERNSTEIN<sup>1</sup>

*Veterans Administration Center, Wood, Wisconsin*

AND RICHARD H. DANA

*West Virginia University*

Dollin and Sakoda have suggested that standard order of TAT administration will produce happier stories toward the end of the series as a function of an adaptation-level effect, and that this effect could be lessened by administering the cards in alternate happy-sad order. To test their prediction, comparisons were made between Cards 1-10 and 11-20 on 67 Ss whose TAT stories had been rated for emotional tone and sadness of outcome. These comparisons yielded *t* ratios, significant beyond the .001 level of confidence, in a direction opposite from the prediction. The ranks of TAT cards for various samples of data were compared for emotional tone and outcomes, and indicate that standard order of TAT administration contains an approximate alternation of happy-sad order. Changing the standard order of TAT administration does not appear warranted on the basis of current evidence.

Dollin and Sakoda (1962) have suggested that, as a consequence of the relative sadness of TAT pictures, an adaptation to sadness occurs when TAT cards are administered in standard order. Thus, happier stories should occur to cards seen late in the series. Evidence concerning this prediction is available from a previous study (Bernstein, 1956), in which TAT cards were administered in a standard order, with four methods of administration. Bernstein's TAT stories had been rated for emotional tone (Eron, Terry, & Callahan, 1950) and for sadness of outcome (Eron, 1950). According to Dollin and Sakoda's prediction, the ratings for outcome and emotional tone should be happier on the last 10 cards than on the first 10.

Table 1 shows *t* test comparisons of ratings on these scales for Cards 1-10 with 11-20, under each method of administration. Two of the four comparisons for emotional tone and three of four comparisons for outcomes are significant in a direction opposite from that predicted by Dollin and Sakoda (1962). The non-significant comparisons show opposite trends; none of the eight comparisons supports their prediction. When the four groups are combined

( $N = 67$ ), the differences between emotional tone and outcome means for the first and last 10 cards are significant beyond the .001 level.

As a means of comparing picture rating with story rating, Dollin and Sakoda (1962) compared the ranks for emotional tone of the 19 cards they used in common with Eron et al. (1950), with a resulting  $\rho$  of .81 for men and .76 for women. A similar comparison of our female subjects with those of Eron et al. (1950) yields a  $\rho$  of .74. When the ranks of the cards for emotional tone for our subjects are compared with Eron's (1953) normative data for women, the resulting  $\rho$  is .76. In comparing the ranks of the cards for outcome for our subjects with those of Eron's (1953) normative data for women, the  $\rho$  is .81. These findings indicate high consistency for the relative ranks of the TAT cards for mood ratings between the samples used by Dollin and Sakoda (1962); Eron et al. (1950); Eron (1953); and Bernstein (1956). However, an inspection of the ranks of the cards for the various samples (Table 2) indicates that there are as many or more sad rankings (1-10) among the last 10 cards as among the first 10.

Dollin and Sakoda (1962) suggest that the adaptation-level effect might be minimized by

<sup>1</sup> Also affiliated with the Marquette University School of Medicine.

TABLE 1  
DIFFERENCES IN MEANS FOR EMOTIONAL TONE AND OUTCOMES  
BETWEEN CARDS 1-10 AND 11-20

Method of administration	Emotional tone				Outcomes		
	N	M(1-10)	M(11-20)	t	M(1-10)	M(11-20)	t
Oral, examiner absent	17	-9.59 <sup>a</sup>	-11.29	1.75	-3.65	-5.65	1.47
Oral, examiner present	16	-4.13	-5.38	1.30	3.69	0.00	3.80*
Written, examiner absent	18	-8.50	-11.33	3.58*	.50	-7.11	7.18*
Written, examiner present	16	-3.25	-5.63	4.67*	4.69	-1.00	3.53*

\* The greater the negative quantity, the sadder the rating for emotional tone and outcomes.  
\*  $p < .01$ .

"noting the stimulus values of the pictures and administering them in alternate happy-sad order [p. 344]." Again, inspection of Table 2 reveals the adequacy of the standard order in meeting their suggestion. In fact, Dollin and Sakoda's data reveal that of the 7 cards rated happiest, 4 are numbered lower than 10 in the standard order; of the 7 cards rated saddest, 4 are numbered higher than 10 in the standard order.

This reanalysis of data from a previous study fails to confirm the Dollin and Sakoda (1962) prediction that TAT cards seen toward the end of the series would evoke happier stories than

those seen earlier. A significant difference in the opposite direction from the prediction was demonstrated. Our findings indicate that in the standard order of administration, the stimulus values of the cards are well randomized.

#### REFERENCES

- BERNSTEIN, L. The examiner as an inhibiting factor in clinical testing. *J. consult. Psychol.*, 1956, 20, 287-290.  
DOLLIN, ADELAIDE, & SAKODA, J. M. The effect of order of presentation on perception of TAT pictures. *J. consult. Psychol.*, 1962, 26, 340-344.  
ERON, L. D. A normative study of the Thematic

TABLE 2  
RANK ORDER OF TAT CARDS FOR EMOTIONAL TONE AND OUTCOMES

Card	Emotional tone			Outcomes	
	Eron (1953) <sup>a</sup>	Eron et al. (1950)	Bernstein (1956)	Eron (1953) <sup>b</sup>	Bernstein (1956)
1	15	11	18.5	15.5	15
2	13.5	12.5	13	17	16
3	2	4	3	3	7
4	6	6	5	12	10
5	12	15.5	8	8	9
6	5	14	16	11	11
7	16	17	9	15.5	13
8	7	19	14	9	17
9	8	5	11.5	5	6
10	18	12.5	20	20	20
11	9	7	6	13	4
12	10	10	11.5	7	8
13	1	1	1	1	2
14	19	20	15	18	18
15	4	3	4	4	1
16	20	18	18.5	19	19
17	13.5	9	10	6	12
18	3	2	2	2	2
19	17	15.5	17	14	14
20	11	8	7	10	5

Note.—The rank of 1 signifies the saddest card; the rank of 20 the least sad of the 20 cards.  
<sup>a</sup> Computed from Table 2, Eron (1953).  
<sup>b</sup> Computed from Table 3, Eron (1953).

Apperception Test. *Psychol. Monogr.*, 1950, 64(9, Whole No 315).

ERON, L. D. Responses of women to the Thematic Apperception Test. *J. consult. Psychol.*, 1953, 17, 269-282.

ERON, L. D., TERRY, DOROTHY, & CALLAHAN, R. The use of rating scales for emotional tone of TAT stories. *J. consult. Psychol.*, 1950, 14, 473-478.

(Received November 9, 1962)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 6, 535-536

## VALIDITY OF THE ROSENZWEIG PICTURE-FRUSTRATION STUDY WITH FELONS AND DELINQUENTS

SAUL ROSENZWEIG

Washington University, St. Louis

In the recent Brief Report by Mercer and Kyriazis (1962) a comparison of a group of physically assaultive prisoners with a group of matched normals on the Rosenzweig Picture-Frustration (P-F) Study yielded no significant differences. This result is not an isolated one. Similar findings for such populations have been reported by Fry (1949), Holzberg and Hahn (1952), and Norman and Kleinfeld (1958). Kaswan, Wasman, and Freedman (1960) in a quite carefully designed and comprehensive investigation got somewhat more favorable results. Moreover, in an investigation by Lindzey and Tejessy (1956) involving both the P-F Study and, more centrally, the TAT, it was found that, contrary to hypothesis, "signs" of aggression on the TAT correlated outstandingly with subjects' self-ratings—not with ratings assigned by a well-informed diagnostic council; and that the TAT variables correlated highly with scores on the P-F. The P-F proved to be at least as good an indicator of aggression as was the TAT.

Since Mercer and Kyriazis in their brief and cautious conclusion state without further elaboration that their results "underscore the difficulty in relating reported attitudes to overt behavior," it appears worthwhile to note some of the interpretive considerations that underlie this difficulty. A technique like the P-F may operate at any one of several (usually) undefined levels—opinion, overt or implicit (Rosenzweig, 1950). These levels must obviously be taken into account in research on delinquents and criminals (Rosenzweig, 1960, p. 171) and, indeed, quite generally. Assaultive delinquents may be well versed in the denial of their hostile tendencies and, if so, would, at the opinion level, obtain normal or even "better than normal" extrapunitive scores. The paradoxical results of Lindzey and Goldwyn (1954) are probably to be so explained. There is also the less obvious possibility that, at the implicit level, some criminals may

actually be more intropunitive than normals (*vide* the psychoanalytic concept of criminality from a sense of guilt); and, if this level is being tapped by the P-F, the scores of such individuals on intropuniveness might actually be higher than those of matched normals. It might then be the exception rather than the rule to find delinquents performing on the P-F at the overt level and thus reflecting the elevated hostility that had placed them in the asocial group.

In short, the problem of determining the level at which a subject performs on any projective device at a particular time remains crucial and unsolved, not only for the P-F Study but for the entire range of these diagnostic instruments. (The P-F has the minor virtue of making this point somewhat more evident than do other projective methods which are more complex and meaningful unless the data also supply a definite context for understanding why the results were negative, i.e., why they were contrary to the explicit or implicit hypothesis under which the work was conducted.

### REFERENCES

- FRY, F. D. A study of reactions to frustration in 236 college students and in 207 inmates of state prisons. *J. Psychol.*, 1949, 28, 427-438.
- HOLZBERG, J. D., & HAHN, F. The Picture-Frustration technique as a measure of hostility and guilt reactions in adolescent psychopaths. *Amer. J. Orthopsychiat.*, 1952, 22, 776-795.
- KASWAN, J., WASMAN, M., & FREEDMAN, L. Z. Aggression and the Picture-Frustration Study. *J. consult. Psychol.*, 1960, 24, 446-452.
- LINDZEY, G., & GOLDWYN, R. M. Validity of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1954, 22, 519-547.
- LINDZEY, G., & TEJESSY, CHARLOTTE. Thematic Apperception Test: Indices of aggression in relation to measures of overt and covert behavior. *Amer. J. Orthopsychiat.*, 1956, 26, 567-576.
- MERCER, MARGARET, & KYRIAZIS, C. Results of the Rosenzweig Picture-Frustration Study for phys-



ically assaultive prisoner mental patients. *J. consult. Psychol.*, 1962, 26, 490.  
NORMAN, R. D., & KLEINFELD, G. J. Rosenzweig Picture-Frustration Study results with minority group juvenile delinquents. *J. genet. Psychol.*, 1958, 92, 61-67.  
ROSENZWEIG, S. Levels of behavior in psychodiagnosis with special reference to the Picture-Frustra-

tion Study. *Amer. J. Orthopsychiat.*, 1950, 20, 63-72.  
ROSENZWEIG, S. The Rosenzweig Picture-Frustration Study, Children's Form. In A. I. Rabin & Mary Haworth (Ed.), *Projective techniques with children*. New York: Grune & Stratton, 1960.

(Received November 13, 1962)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 6, 536-537

NOTE ON TESTING CONDITIONS<sup>1</sup>

ROBERT C. PIERCE

*Langley Porter Neuropsychiatric Institute, San Francisco*

Authors of intelligence tests usually recommend that their test be administered in quiet, undistracting surroundings. Wechsler (1955), for example, says that testing should be done in a room "free from distracting noises and intrusions."

As part of a large multidisciplinary study of geriatric mental illness (see Crook & Katz, 1962; Simon & Fiske, 1960), we have had occasion to administer four Wechsler Adult Intelligence Scale (WAIS) subtests (Information, Comprehension, Arithmetic, Digit Span) to several hundred patients over 60 years of age admitted to the psychiatric screening ward of San Francisco General Hospital. Patients were tested upon admission to the ward and approximately a year and 2 years later. At the time of the first and second retests, the interviewer administering the test made a rating of the conditions under which testing took place; the rating was a simple 3-point scale of "good,"

"fair," and "poor" based on the frequency and disruptiveness of interruptions, noise, and intrusions. A rating of good was given if the testing was completed in a continuous session with no more than one or two minor noises or distractions, such as traffic noise or the voices of neighbors. A rating of poor was given if there were several distracting noises or if one or more very disruptive events occurred (such as "kibitzing" of fellow patients in a rest home).

The unit of analysis is the sum of the scaled scores for the four subtests. Table 1 presents the means and standard deviations of WAIS scores for patients tested under good, fair, and poor conditions at first and second retests. One-way analyses of variance indicate that means for the three testing conditions do not differ at either first retest or second retest.

Since there were several interviewers (16 during the first retest, 9 during the second), it is possible that differences in criteria for rating testing conditions account for the absence of the expected relationship between score and testing conditions. Separate one-way analyses of variance were therefore computed for each interviewer who tested more than 20 patients. None of the

<sup>1</sup> This research was supported by grants from the National Institute of Mental Health (3M9145(C4)) and the California Department of Mental Hygiene (62-1-16.3).

TABLE 1  
WAIS BY TESTING CONDITIONS, HOSPITAL SAMPLE

Testing condition	First retest			Second retest		
	N	Mean WAIS <sup>a</sup>	SD	N	Mean WAIS <sup>a</sup>	SD
Good	50	30.1	11.4	107	27.5	12.5
Fair	9	27.6	8.0	40	24.8	10.9
Poor	2	28.6	5.5	2	28.5	1.5 <sup>b</sup>
		$F_{2,62} = .246$			$F_{2,148} = .744$	

<sup>a</sup> Each individual's score is the sum of scaled scores for Information, Comprehension, Arithmetic, and Digit Span subtests. The mean of the individual scores is shown here.  
<sup>b</sup> When fair and poor are combined and tested against good,  $t = 1.16$ ,  $p > .10$ .

TABLE 2  
WAIS BY TESTING CONDITIONS,  
COMMUNITY SAMPLE

Testing condition	N	Mean WAIS <sup>a</sup>	SD
Good	236	39.3	10.1
Fair	27	37.6	10.0
Poor	2	34.0	12.0
$F_{2,262} = .604$			

<sup>a</sup> Each individual's score is the sum of scaled scores for Information, Comprehension, Arithmetic, and Digit Span subtests. The mean of the individual scores is shown here.

analyses revealed significant differences between the three testing conditions.

Although unlikely, it is also possible that the large variability of scores found in a mentally ill sample obscured small but real differences among the testing conditions. Such differences might be revealed by examining a normal sample. Another phase of the project called for testing 265 people over 60 years of age who were living in San Francisco. The same 3-point rating of testing conditions was available. Table 2 presents a one-way analysis of variance on the means for the three testing conditions. Again, no significant differences are found.

In view of the widespread assumption that testing conditions have an important effect upon test scores, it is surprising to find that such an effect does not appear under the extreme range of conditions we encountered. Our rating of testing conditions is admittedly a crude one and will allow us to conclude only that poor testing conditions do not have a gross effect upon test scores. A more homogeneous group, such as college sophomores, might be better suited to a test of the hypothesis that distractions and noise affect test results, but at the least we have evidence that intelligence tests are more robust than test authors have thought.

#### REFERENCES

- CROOK, G. H., & KATZ, L. Patterns of geriatric mental illness: Intellectual functioning of aged patients and non-patients. In C. Tibbetts & Wilma Donahue (Eds.), *Social and psychological aspects of aging*. New York: Columbia Univer. Press, 1962.
- SIMON, A., & FISKE, MARJORIE. Studies in geriatric mental illness: Some research plans and assumptions. *J. Amer. Geriat. Soc.*, 1960, 8, 363-372.
- WECHSLER, D., *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

(Received December 3, 1962)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 6, 537-539

## INVESTIGATION OF THE STABILITY OF COLOR PREFERENCES

WALTER A. HAAS

*Veterans Administration Hospital, Lyons, New Jersey*

An experiment to determine color preferences was performed with 20 college students who were their own controls. 3 techniques were used: tiles, free painting, and questionnaire. Relationships between the 3 techniques were positive and significant at the .01 level. The stability of the color choices was investigated by repeating the same 3 techniques 4 weeks later. Data indicated that the color preferences remained consistent. Significance was at the .01 level. Ss were exposed to stimuli which produced frustration or success to note whether the color preferences could be altered. Data obtained with the chi square method indicated that the established color preferences were not altered by these stimuli.

Past investigations have dealt with the color preferences of specific groups; for example,

<sup>1</sup> This paper is adapted from a dissertation submitted to New York University in partial fulfillment of the requirements for the PhD degree. The writer wishes to acknowledge the invaluable assistance of the sponsoring committee, Edward Kemp, H. Harry Siles, Hale Woodruff, and Florence Halpern.

psychotic patients, Indians, Negroes, infants (Block & Caldwell, 1959; Garth, 1922; Mercer, 1925; Staples, 1932). A number of studies have connected specific personality attributes to choice of specific colors. Napoli (1951) associated green with creativity. Naumburg (1953) ascribed hate and depression to the color black. Alschuler and Hattwick (1947); Brick (1944); and

TABLE 1  
SUMMARY OF PROCEDURE

Sitting 1	Sitting 2
Part A	Part A
1. Tile choices	Steps 1, 2, 3, of Sitting 1 Repeated
2. Free painting	
3. Questionnaire	
Part B	Part B
1. Frustration Stimuli (word reconstruction)	1. Success stimuli (word reconstruction)
2. Three designs (tiles)	Steps 2, 3 of Sitting 1 repeated
3. Three designs (free painting)	

Note.—Sitting 2 was presented 4 weeks after Sitting 1.

Lehman and Risquez (1953) have inferred similar relationships.

The present study is concerned with the determination of color preference, the consistency thereof, and the effect of manipulation on the established color preference base.

### METHOD

#### Subjects

Twenty undergraduate students from the University of Delaware volunteered for the study. Since each subject was seen individually and since he was his own control, no further delimitations on the selection of subjects were imposed.

#### Procedure

Subjects manifesting color blindness were excluded by the Jensen test for color blindness. Color preferences were obtained by three methods: tiles, free painting technique, and a questionnaire. To note consistency of the given color preferences, the three methods were repeated at least 4 weeks later.

**Tiles.** Six tiles (red, blue, green, black, white, yellow) were taken from the Lowenfeld (1952) mosaic test. Each subject was asked to arrange the tiles in the order of his preference.

**Free painting technique.** Tempera paint was chosen. To facilitate a comparison, the same colors were selected. Gray paper was used so that each color, especially white, could show up sufficiently well. The subject was told: choose a color and paint anything you like. The subject could use only one color per paper. After a particular color was used, it was taken away. The instructions were repeated until all colors were used.

**Questionnaire.** Color choices were obtained by the questionnaire without presentation of visual stimuli. Questions pertaining to the rationale of the order of the colored tiles, factors that influenced the selection of particular color paints and their subsequent productions yielded qualitative data.

The summary of the procedure is given in Table 1.

Stimuli which produced frustration and success were utilized to note whether the color preference base could be altered. The object of the frustration stimuli was to interfere, interrupt, or completely prevent the completion of goal-directed responses. The aim of the success stimuli was to produce a favorable and desirable termination of a goal-directed response. The frustration stimuli were given during the first session and the success stimuli were administered 4 weeks later.

**Frustration Stimuli.** Seventy words from a New York City school principal examination were chosen. These words were unusual, difficult, and complex. The letters of each word were so arranged to prevent recognition. To illustrate: aacentttu (attenuate), acehnopt (cenotaph), acerrst (serrate). To maintain the subject's motivation every seventh word on the list was an easy word. A total time limit of 50 minutes was imposed and the time per word was limited to 40 seconds. During the administration, such disparaging remarks as, "For a college student you are doing poorly" were made. These were the written instructions: "Rearrange the letters to make a proper word." Below the instructions three disarranged simple words provided a sample.

**Success Stimuli.** Forty easy words from a seventh grade public school test were selected. To illustrate: omtor (motor), ciily (civil), conab (bacon). A total time limit of 50 minutes was imposed. The performance of the subject was encouraged by such remarks as, "You are doing fine." The written instructions were: "Rearrange the letters to make a proper word."

Below the instructions three disarranged simple words served as a sample.

After the presentation of the frustration stimuli and the success stimuli, the subjects were asked (a) to make three designs with tiles, using only one color per design (the entire mosaic test was utilized for this purpose) and (b) to make three paintings choosing one color, per paper. The same color could have been repeated, as no order of choice was requested. The object was to study responses to the frustration and success stimuli.

### RESULTS AND DISCUSSION

The relative standings of each subject on the three different techniques was examined with the Spearman rank correlation method. The data in Table 2 indicate positive correlations between the responses obtained by the three techniques. All correlation coefficients were significant at the .01 level. Correlation between color preferences obtained by the tiles and the questionnaire were



TABLE 2

COMPARISON OF COLOR PREFERENCES AMONG THE  
THREE TECHNIQUES EXPRESSED IN  
RANK CORRELATIONS

	Tiles with free painting	Tiles with question naire	Free painting with question naire
Sitting 1	.537	.703	.518
Sitting 2	.655	.809	.548

higher than those between the color preferences obtained by the free painting technique and the other two techniques. There were qualitative suggestions that form and content played a more important role on the free painting technique. Color preferences remained reliable over the 4 week period (tiles  $\rho = +.739$ , free painting technique  $\rho = +.567$ , questionnaire  $\rho = +.779$ ). All correlations were significant at the .01 level.

The hypothesis that the color preference base can be manipulated by frustration and success stimuli is rejected. The order of preference was maintained. The first three places of the color preference base were used in 71.7% of the tile designs and paintings. Significance was investigated with the chi square method and found to be at the .01 level. Data indicate that the stimuli used in an effort to manipulate the color preference base served the desired objectives. Remarks, performance, and overt behavior of the subjects during the administration of the

frustration stimuli indicate that responses to the tasks were interfered with, interrupted, and incomplete. Success stimuli on the other hand were easy and readily dealt with.

Results of the study support the conclusion that color preferences remained stable over the 4-week period and were not affected by the frustration and success stimuli.

## REFERENCES

- ALSCHULER, R. H., & HATTWICK, L. A. *Painting personality*. Chicago: Chicago Univer. Press, 1947. 2 vols.
- BLOCK, D. S., & CALDWELL, W. Relationship between color and affect in paranoid schizophrenics. *J. gen. Psychol.*, 1959, 61, 231-255.
- BRICK, M. Mental hygiene value in children's art work. *Amer. J. Orthopsychiat.*, 1944, 14, 136-144.
- GARTH, T. R. A color preference of 559 full-blood Indians. *J. exp. Psychol.*, 1922, 5, 392-418.
- LEHMAN, H., & RISQUEV, F. A. The use of finger painting in the clinical evaluation of psychotic conditions: A qualitative and quantitative approach. *J. ment. Sci.*, 1953, 99, 763-777.
- LOWENFELD, MARGARET. *The Lowenfeld mosaic test*. New York: Grune & Stratton, 1952.
- MERCER, F. M. Color preferences of 1006 Negroes. *J. comp. Psychol.*, 1925, 5, 109-146.
- NAPOLI, P. J. Finger painting. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective psychology*. New York: Prentice Hall, 1951, 386-415.
- NAUMBURG, M. *Psychoneurotic art: Its function in psychotherapy*. New York: Grune & Stratton, 1953.
- STAPLES, R. The responses of infants to colors. *J. exp. Psychol.*, 1932, 15, 119-141.

(Received January 16, 1963)

## THE PROBLEM OF SELECTING DRUG-FREE SCHIZOPHRENICS FOR RESEARCH<sup>1</sup>

LOREN J. CHAPMAN

*Southern Illinois University*

In selecting patients to be removed from phenothiazine therapy for purposes of research, 38 chronic male schizophrenics at a VA hospital were given 1 of 2 screening tests. These were paper-and-pencil measures of conceptual disorder. Before a 6-week "drying-out" period had been completed, 18 of these patients showed an accentuation of their psychotic symptoms, and so were removed from the study and returned to medication by the hospital staff. These 18 patients were found to have shown on the screening tests a significantly greater disorganization of thought than the remaining 20. This points out a hazard for investigators who must rely for their research on patients who do not profit from tranquilizing drugs. Findings based on such patients may be limited in their generality.

With the widespread use of tranquilizing drugs, it is frequently difficult to obtain drug-free samples of schizophrenics for purposes of research. Investigators often select their patients without regard to their current drug status, perhaps with the rationalization that it is not yet known just what effect the drugs have anyway. This is a perilous procedure, since the drugs clearly influence some aspects of the schizophrenic psychosis in many patients, and if the investigator's measure is tapping something of relevance to schizophrenia, there is a good chance that differences between normals and schizophrenics may be washed out by the drug.

Another alternative is to test only those patients who are not currently receiving such drugs. In most hospitals, however, this group is not only small, but it is also likely to be comprised primarily of patients who show few overt psychotic symptoms—hardly a suitable group with which to study schizophrenia.

A third alternative, the one of concern in the present paper, is to ask the hospital administration to remove from drugs a group of schizophrenics for purposes of research. Before the patients are actually tested they must go through a "drying-out" period, which allows the drugs to be eliminated from the body. However, many of the patients may then show an accentuation of their psychotic symptoms. The

staff are likely to become concerned for the patients' welfare and to ask that particular patients be returned to medication. This may leave the investigator with an atypical schizophrenic sample, namely, those for whom the withdrawal of drugs does not result in an increase of symptoms. This group may differ on psychological measures from the remaining patients who are helped by drugs. The present report presents some evidence on this point.

Several investigators arranged at a Veterans Administration hospital to withdraw patients from drugs for studies which required drug-free schizophrenics. The subjects were white, male, long-term, chronic schizophrenics, between 21 and 56 years of age, with at least 6 years education, whose most recent admission was no more recent than 6 months, who showed no signs of mental deficiency or brain damage, and who had not received electric shock therapy in the previous 6 months. Patients were also selected by ability to take one of two simple paper-and-pencil screening tests, like those to be used in the research.

### METHOD

*Instruments.* The screening tests were a homonym test<sup>2</sup> (Burstein, 1961) and a paper-and-pencil version of a breadth of concept test (Chapman & Taylor, 1957). The homonym test consisted of 48 items of the following format:

Circle the word that means almost the same as sea.

See                      Ocean                      Idol

The correct answer, of course, is "Ocean." The tendency to confuse homonyms with synonyms would be reflected by the choice of the "See"

<sup>2</sup> The author is indebted to Alvin Burstein for supplying a copy of his homonym test.

<sup>1</sup> This study was supported by a research grant MH-3481 from the National Institute of Mental Health, United States Public Health Service. The author is indebted to the Veterans Administration Hospital in Lexington, Kentucky for permission to test patients. Glenn Miller assisted in testing the patients and in analyzing the data.

TABLE 1

MEAN ERROR SCORES OF PATIENTS SUBSEQUENTLY KEPT OFF DRUGS AND RETURNED TO DRUGS

	Homonym test			Breadth of concept test			
	Homonym	Unidentified error	Total	Inclusion similar	Inclusion dissimilar	Exclusion	Total
Kept off drugs	2.4	.6	3.0	7.4	.4	3.4	11.2
Returned to drugs	17.5	1.0	18.5	10.7	1.6	7.7	20.0

alternative. The third alternative, "Idol," is intended to bear no relationship to the stimulus word, and is included in the item as a measure of errors of unidentified origin. For a further description of this test, see Burstein (1961).

The breadth of concept test was presented in the following format:

Circle each word which names a fruit.

Cross out each word which does *not* name a fruit.

Peach Screwdriver Saw Peas Corn Pear, (etc.)

The breadth of concept test contained six tasks like the one above, each of which had 30 items which were to be marked as belonging, or not belonging, to the named concept, and in each case the 30 alternatives included 10 items which correctly belonged in the concept (e.g., "fruit," above), 10 which did not belong but were similar to the correct items (e.g., "vegetables," above), and 10 which did not belong and were dissimilar (e.g., "tools," above). The six named categories for the tasks were "fruit," "trees," "alcoholic drinks," "animals found in the home or on the farm," "kitchen tools or implements," and "articles of clothing." For a further description of this test, see Chapman and Taylor (1957).

Half the subjects were given the homonym test for screening purposes, and half were given the breadth of concept test. All of these were scheduled to be removed from medication for 8 weeks, the first 6 weeks of which were to be used as a drying-out period, and the last 2 weeks for testing.

**Subjects.** Of the 111 patients who were seen, 52 took one of the two screening tests successfully. The hospital management consented to remove 38 of them from medication, explaining that 12 of the remaining 14 were too disturbed, and two were about to leave the hospital. Of these 38 patients, 20 had taken the homonym test, and 18 the breadth of concept test. Their mean age was 39.7, and mean education was 10.8. All of these patients were receiving phenothiazine therapy. 12 were receiving chlorpromazine, dosages ranging from 200 to 800 milligrams a day. Others were receiving trifluoperazine, prochlorperazine, and thioridazine.

After the patients were removed from medication, many of them began to show an exacerbation of

their psychotic symptoms. Some became more withdrawn, or untidy, or openly delusional, or showed ritualistic gestures. The nursing staff expressed concern to the ward physicians, and they in turn asked the hospital administration to remove certain patients from the research project. This continued until 18 of the 38 patients were returned to medication. These 18 patients were returned after a mean of 26.9 drug-free days.

When the experimental testing was begun after the 6 week drying-out period, it was noticed that a surprisingly large proportion of the patients seemed to be performing in a fairly normal manner. In order to get some light on this, these patients were compared as to their earlier performance on the screening tests with those patients who had been prematurely returned to medication. At the time of the screening tests, all patients were, of course, on drugs. Of the patients prematurely returned to drugs, 11 had taken the homonym test, and 7 the breadth of concept test. Of the patients maintained off drugs, 9 had taken the homonym test, and 11 the breadth of concept test.

## RESULTS

The findings are shown in Table 1. As seen there, the two groups appeared to differ greatly in their initial pathology as reflected by performance on the screening tests. On both the homonym and the breadth of concepts tests, the patients who later were returned prematurely to drugs made more total errors. This was significant at the 5% level in each case as determined by a two-tailed Mann-Whitney U test. On the homonym test, the difference score of homonym errors minus unidentified errors also distinguished the two groups at the 5% level of significance.

As the result of these findings, the investigators decided that they could not rely exclusively on the patients who were maintained off drugs as a source of data concerning the nature of schizophrenic psychopathology. They obtained elsewhere another sample of patients who were kept off drugs regardless of their symptoms.



## CONCLUSIONS

Investigators who compare groups of schizophrenics with normals should take into account the drug status of their patients. In addition, if they use patients who were not currently receiving drugs, they should attempt to obtain a sample which is not biased by selective exclusion of subjects from the sample. Any conclusions that are reached using only patients whose symptoms do not worsen with the withdrawal of tranquilizing drugs may be limited in their generality.

Schizophrenics who profit from phenothiazine

therapy appear to differ in their thought disorder from patients who do not. This might be worth further exploration by investigators who are concerned with finding subvarieties of schizophrenia.

## REFERENCES

- BURSTEIN, A. G. Some verbal aspects of primary process thought in schizophrenia. *J. abnorm. soc. Psychol.*, 1961, 62, 155-157.  
 CHAPMAN, L. J., & TAYLOR, JANET A. Breadth of deviate concepts used by schizophrenics. *J. abnorm. soc. Psychol.*, 1957, 54, 118-123.

(Received January 25, 1963)

*Journal of Consulting Psychology*  
 1963, Vol. 27, No. 6, 542-544

## THE RELATION BETWEEN MATERNAL SELF-ACCEPTANCE AND CHILD ACCEPTANCE<sup>1</sup>

GENE R. MEDINNUS AND FLOYD J. CURTIS

*San Jose State College*

2 measures of self-acceptance and one measure of child acceptance were administered to 56 mothers of children enrolled in a parent-participating cooperative nursery school in order to test the hypothesis of a significant positive relation between maternal self-acceptance and child acceptance. The hypothesis was confirmed. The semantic differential technique for assessing child acceptance holds promise for further research in this area.

Assessment of parental attitudes has received a tremendous amount of attention in research concerned with factors influencing the child's personality development. While the content of most current parent attitude tests focuses on the parent's attitude toward a number of aspects of child rearing, of at least equal importance is the parent's attitude toward the individual child in terms of an acceptance-rejection dimension. In case studies of emotionally disturbed children, the inference frequently is made that one or both of the parents reject the child. Support for the emphasis on the importance of parental acceptance comes also from factor analyses of parent behavior ratings (Schaefer, 1959) and of parent attitude scales (Zuckerman, Ribback, Monashkin, & Norton, 1958). In both analyses, an acceptance-rejection factor appeared in addition to an authoritarian-democratic one.

A number of personality theorists have suggested that the attitudes held toward the self

<sup>1</sup> A summary report of this study was presented at the biennial meeting of the Society for Research in Child Development, Berkeley, California, April, 1963.

are reflected in the attitudes held toward others. Fromm (1939), for example, maintains that self-love and the love of others go hand in hand. He notes that a failure to love the self is accompanied by a basic hostility toward others. Horney (1950) feels that the person who does not love himself is incapable of loving others. Rogers (1951) argues that when a client enters therapy he typically holds a negative self-concept; he is unable to accept himself. However, once the client becomes more accepting of himself, he begins to be more accepting of others.

The attitudes which a child holds toward himself, especially those dealing with self-esteem and self-worth, play an important role in his personality development. The extent to which a child develops a positive self-concept depends crucially upon the extent to which he is accepted by the "significant others" (typically his parents) in the early years.

While numerous studies could be cited showing a relation between self-acceptance and acceptance of others, a majority of these have used as subjects college students or individuals receiving counseling or therapy. There is almost no re-

TABLE 1  
CORRELATIONS AMONG MATERNAL SELF-ACCEPTANCE AND CHILD ACCEPTANCE MEASURES

Measures	Bills self-acceptance	Bills adjustment	Semantic differential child acceptance
Semantic differential self-acceptance	-.57**	.53**	.33*
Bills self-acceptance		-.57**	-.48*
Bills adjustment			.28

\*  $p < .05$ .

\*\*  $p < .01$ .

search evidence for the relation between parental self-acceptance and acceptance of the child, though several investigations have described the effects on the child's personality development of parental nonacceptance. Most of this research has dealt with clinic populations, however, and the research involving normal parents and children is meager and less confirmatory.

#### HYPOTHESIS

The purpose of the present investigation was to test the hypothesis that there is a significant positive relation between self-acceptance and child acceptance in a nonclinical group of mothers of young children.

#### METHOD

##### *Subjects*

Fifty-six mothers of children enrolled in a parent-participating cooperative nursery school served as subjects. The mothers ranged in age from 25 to 45 years with a mean of 32; years of education ranged from 11 to 18 with a mean of 14. The children in the nursery school on whom the mothers made the ratings ranged in age from 3 to 5 with a mean of 4 years.

##### *Procedure*

Two measures of self-acceptance and one measure of child acceptance were administered during two consecutive meetings of the mother group. The self-acceptance measures were the Bills Index of Adjustment and Values (Bills, Vance, & McLean, 1951) and a semantic differential scale of 20 bipolar adjectives in which the distance (D) between the mothers' ratings of "Me (as I am)" and "Me (as I would most like to be)" was defined operationally as the extent of self-acceptance. The child acceptance measure consisted of the same set of bipolar adjectives with the distance between the mothers' ratings of "My child (as he is)" and "My child (as I would most like him to be)" defined as the extent of maternal acceptance of the child.

#### RESULTS AND DISCUSSION

The intercorrelations among the three self-regard scores and the child acceptance measure are provided in Table 1. All of the coefficients are in the expected direction; the negative correlations are due to the method of scoring the tests. On the Bills self-acceptance measure, the higher scores indicate greater self-acceptance while the Bills adjustment scores and the semantic differential self-acceptance scores are discrepancy scores with the higher scores denoting less favorable self-attitudes. The findings strongly support the hypothesis of a significant positive relation between maternal self-acceptance and child acceptance.

It is likely that the present sample of mothers was restricted in terms of the child acceptance dimension. The mothers were concerned and interested enough in their children to spend the considerable time and effort which is involved in participation in the cooperative nursery school. Since moderate correlations were obtained in this rather restricted range of subjects, one might expect higher correlations to emerge in a study of mothers randomly selected from the general population.

The validity of the measure employed to assess child acceptance remains to be established. Subsequent research may show that factors in the parent-child relationship other than acceptance are reflected in this instrument. However, the fact that positive results were obtained in the present study suggests that the semantic differential approach holds promise as an instrument for assessing the dimension of parental acceptance of the child. The technique is easy to administer and score, and its purpose in assessing child acceptance appears to be somewhat disguised.

The present findings suggest that the area of parental self-acceptance merits intensive study in any attempt to identify antecedents of child acceptance, an important element in the child's self-acceptance and in his personality development in general.

## REFERENCES

- BILLS, R. E., VANCE, E. L., & MCLEAN, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, 15, 257-261.
- FROMM, E. Selfishness and self-love. *Psychiatry*, 1939, 2, 507-523.

- HORNEY, KAREN. *Neurosis and human growth*. New York: Norton, 1950.
- ROGERS, C. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
- SCHAEFER, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.*, 1959, 59, 226-235.
- ZUCKERMAN, M., BARRETT-RIBBACK, BEATRICE, MONASHKIN, I., & NORTON, J. Normative data and factor analysis on the parental attitude research instrument. *J. consult. Psychol.*, 1958, 22, 165-171.

(Received February 18, 1963)

*Journal of Consulting Psychology*  
1963, Vol. 27, No. 6, 544-545

# RORSCHACH Z SCORES ON DISTURBED SUBJECTS

JOHN NICKOLS<sup>1</sup>

*Hillcrest Residential Services for Children, Charleston, West Virginia*

Observations on the Rorschach responses of 12 disturbed children and 75 hospitalized neuropsychiatric adults provided evidence that sum Z was significantly related to "breadth in mental efficiency." In this sense, sum Z appeared strongly influenced by (a) the number of mental functions which might be elicited by a task situation, and (b) how efficiently the individual utilizes these functions. The media of communicating responses also seems to relate sum Z to verbal intelligence. The level of Z capacity served as the more useful index in estimating breadth of mental efficiency in children.

The responses of disturbed Ss were observed for empirical evidence relating to the following propositions. (a) If Z denotes a "more accurate" measure of intelligence which is independent of the type of intelligence used (Beck, 1949), there should be no discernible pattern to the correlations between Z and scores obtained on a variety of IQ tests. (b) Although its relation to organization activity (Beck, 1949) suggests that Z might correlate well with scores on perceptual organization, the requirement that scorable Z activity be communicated would also suggest that it might correlate well with a verbal factor.

## METHOD

First, a single examiner tested 12 disturbed children (M age =  $13 \pm 2$ ). MA-equivalents were determined on the WISC (M =  $11 \pm 2$ ), Leiter International Intelligence Scale (M =  $9 \pm 2$ ), and Wide Range Achievement Test (M =  $9 \pm 2$ ). MAs on the Ammons Full Range Picture Vocabulary (M =  $11 \pm 2$ ), Benton Visual Retention (M =  $10 \pm$

2, using revised scoring method; cf., Nickols, 1963a), and Draw-A-Person Tests (M =  $10 \pm 2$ ) were also averaged as Composite scores (M =  $10 \pm 2$ ).

Second, test records were selected randomly on 75 relatively newly hospitalized veterans (M age =  $35 \pm 10$ ). Most of these Ss were diagnosed schizophrenic, but no attempt was made to control for diagnosis because Levine and Spivack (1959) found similarly significant correlations between Rorschach variables on intelligence and Wechsler IQ for hospitalized schizophrenics and nonschizophrenics. Neither was short-term or moderate use of tranquilizers viewed as a significant factor when the Rorschach and Wechsler Scales are employed as the evaluative criteria in research (cf., Nickols, 1958, 1959, & 1962).

For each adult, the WAIS (M IQ =  $91 \pm 18$ ) and Rorschach (M sum Z =  $20 \pm 14$ ) had been administered by the same examiner trained in the Beck system exclusively. WAIS variables were determined on the basis of empirical evidence contributed by Cohen (1959). Perceptual Organization was measured by the sum of the weighted scores on the Block Design and Object Assembly subtests (M =  $17 \pm 6$ ). Sum weighted scores on the Information, Comprehension, Similarities, and Vocabulary subtests quantified a verbal factor of intelligence,

<sup>1</sup>42 Arcturus Lane, Alexandria, Va. Data also from Veterans Administration Hospital, Lexington, Kentucky.



verbally retained knowledge, and the WAIS G factor all in one score ( $M = 35 \pm 13$ ).

### RESULTS

The Rorschach responses of each child were scored according to Beck's system for obtaining sum  $Z$  on the individual protocol. Very significant Rhos (.66 to .72) occurred between (a) sum  $Z$  ( $M = 18 \pm 12$ ), and (b) MAs on the Leiter, Composite, and WISC Scales, in that order. The three sets of MA values had inter-correlations of .67. Sum  $Z$  correlated significantly (about .55) with the Full Range and Draw-A-Person; these latter test scores correlated about .50 with Leiter and WISC scores, in each case. Sum  $Z$  correlated .50 with the Wide Range; .29 with age; .28 with Benton scores.

Results on adults showed sum  $Z$  to correlate .54 with total sum weighted scores for the subtests of the WAIS ( $N = 50$ );  $-.02$  with age ( $N = 75$ ). Sum  $Z$  correlated .43 with verbal activity, and .21 with the perceptual organization factor ( $N = 75$ ); with an  $r$  of .58 between the two  $r$ 's, the difference was significant ( $t_d = 2.3$ ). The greater correlation with the sum of the weighted scores on the WAIS Full Scale would suggest that sum  $Z$  is a quantitative index most significantly related to composite tests capable of eliciting a wide range of mental activity (cf., Nickols, 1963b).

A pattern thus evolved in which sum  $Z$  tended to vary with sampling "breadth" in mental efficiency. A second hypothesis is that sum  $Z$  correlated significantly better with the WAIS Verbal factor. If sum  $Z$  is a measure of thinking power, perhaps it refers more to verbal thought than perceptual organization as an expression of intelligence used to estimate organizational activity. The preferred definition of sum  $Z$ , however, might concern more expressions of mental activ-

ity than those subsumed by verbal abilities alone.

The individual protocols were inspected, and the three responses with the highest  $Z$  score values were used to determine  $Z$  capacity. These sums of the three highest  $Z$  scores obtained on each  $S$  correlated .90 with sum  $Z$  for children; .76 for adults. Intertest correlations between  $Z$  capacity and scores on the six mental tests administered to children, as well as on the WAIS factors determined on adults, formed the same general patterns as obtained with the use of sum  $Z$ . A partial exception was that the size of the coefficients obtained with  $Z$  capacity was noticeably larger for children. Thus,  $Z$  capacity might provide an alternate quantitative expression of intelligence on disturbed children.

### REFERENCES

- BECK, S. *Rorschach's test*. New York: Grune & Stratton, 1949.
- COHEN, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *J. consult. Psychol.*, 1959, 23, 285-299.
- LEVINE, M., & SPIVACK, G. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1959, 23, 306-312.
- NICKOLS, J. A controlled exploratory investigation into the effects of Thorazine upon mental test scores of chronic hospitalized schizophrenics. *Psychol. Rec.*, 1958, 8, 67-76.
- NICKOLS, J. Mental deficit, schizophrenia and the Benton Test. *J. nerv. ment. Dis.*, 1963, 136, 279-286. (a)
- NICKOLS, J. Speculations on the pain variable in psychological research with Thorazine. *Psychol. News.*, 1959, 10, 268-275.
- NICKOLS, J. Structural efficiency of WAIS subtests. *J. clin. Psychol.*, 1963, 19, 420-423. (b)
- NICKOLS, J. Tranquilizers and experimental psychopathology. *Amer. Psychologist*, 1962, 17, 263-264.

(Received February 28, 1963)

## BRIEF REPORTS

### RATER RELIABILITY AND PREDICTION OF DIAGNOSIS WITH THE WITTENBORN PSYCHIATRIC RATING SCALES<sup>1</sup>

DANIEL F. MACHIR<sup>2</sup>

AND

DAVID L. RUSSELL

United States Naval Hospital, Bethesda, Maryland

Ohio University

One chief difficulty of the Wittenborn Psychiatric Rating Scales (WPRS) (Wittenborn, 1955) has been its rather low interrater agreement when used by hospital staff members, including physicians and nurses. Internal consistency reliability measures have not been considered appropriate for this kind of instrument. Since these scales present easily obtainable data for making evaluations and prognoses, the present authors tried to improve interrater agreement by specifying the type and level of training of the raters, and by obtaining ratings under standard conditions. To determine the accuracy of predictions of later hospital staff diagnoses was a further purpose of the study.

In this study 24 mental patients were rated by four staff psychologists of similar training and experience. Each subject was given a 30- to 45-minute structured, fact finding interview, all by the same rater, while being observed through a one-way window by the other three raters. Ratings of each subject on the WPRS completed independently by each observer were then tested for interrater agreement and compared with subsequent psychiatric diagnoses of the subjects.

The results were: (a) Agreement between observers in rating each of nine WPRS symptom clusters tested by Kendall's coefficient of concordance  $W = .86$ , median  $W = .72$ , all significant at the

<sup>1</sup> An extended report of this study may be obtained without charge from Daniel F. Machir, 719 Monroe Street, Rockville, Maryland, or for a fee from the American Documentation Institute. Order Document No. 7631 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

The authors wish to thank L. Marshall, C. Semans, and L. D. Hannah for acting as raters, and H. Fockler for his assistance and cooperation.

<sup>2</sup> Formerly at Athens State Hospital, Athens, Ohio. The statements and opinions contained herein are those of the authors and are not to be construed as official or reflecting the views of the Navy Department or of the Naval Service at large.

TABLE 1

COMPARISONS OF WITTENBORN SCALE CLASSIFICATIONS  
WITH STAFF DIAGNOSES

Staff diagnosis	Wittenborn scale classification				<i>p</i>
	Negative		Positive		
	False	True	False	True	
Psychotic	5	6	1	12	.018
Neurotic	0	17	4	3	.017
"Other"	1	19	1	3	.008

cordance ( $W$ ) produced  $W$ 's ranging from .62 to .91, all significant at the .001 level, thus indicating significant interrater agreement on all WPRS symptom clusters. (b) Comparisons of WPRS composite indexes, classified broadly as Psychosis, Psychoneurosis, and "Other," with subsequent psychiatric diagnoses (uncontaminated), were found significant beyond the .02 level for each major category. Eighteen diagnoses were correctly predicted by the WPRS, or 75% "hits"; six were false positive.

These findings indicate that significant interrater agreement can be obtained using the WPRS under controlled conditions. Further, WPRS diagnostic signs derived from ratings so obtained predicted significantly later broad psychiatric classifications of Psychosis, Psychoneurosis, and "Other," suggesting that WPRS ratings obtained under controlled conditions might be precise enough for clinical diagnostic, as well as descriptive, purposes. Pending further cross validation and refinement of diagnostic indicators, it is suggested that WPRS ratings based upon controlled interview and observational situations might predict broad psychiatric classifications dependably, thus providing a valuable aid to clinical diagnosis.

#### REFERENCE

WITTENBORN, J. R. *Psychiatric Rating Scales*. New York: Psychological Corporation, 1955.

(Received April 2, 1963)

## FACTORS ASSOCIATED WITH INTERRATER DISCREPANCIES ON A PSYCHIATRIC RATING SCALE<sup>1</sup>

ALLEN RASKIN AND PAUL D. SULLIVAN

*Georgetown University*

The Inpatient Multidimensional Psychiatric Scale (IMPS) has been widely used to measure change in patients undergoing various treatment procedures. This instrument was designed for use by psychiatrists and psychologists following a 30- to 60-minute interview. As clinical skills and experience play some part in rating presence and magnitude of specified forms of psychopathology, a portion of the error variance in estimates of interrater reliability may be due to differences between raters on certain background and other factors. Identification of these factors would enhance the reliability of IMPS ratings.

The study sample comprised 51 male and 36 female schizophrenic patients drawn from two large receiving hospitals, four state hospitals, and three private institutions.

To obtain relatively stable measures of interrater discrepancies in IMPS ratings, pairs of raters were selected who had jointly seen at least three different patients at either the pretreatment period or at the end of the patient's sixth week on medication. There were 29 pretreatment and 26 six-week rater pairs. Eight rater variables sampled the rater's clinical experience, experience with personal therapy, weekly time spent with each patient, contact with patient's relatives or friends, frequency of discussion of patient's dynamics, symptoms and problems with the other

IMPS rater, time to complete IMPS, opinion of efficacy of tranquilizers and age. A rational scheme was developed for ordering interrater dissimilarity on seven of these variables. Age difference was not categorized.

Ebel's analysis of variance technique for estimating the reliability of ratings was used to compute the measures of interrater dissimilarity in level of ratings (Level Difference) and dissimilarity due to nonsystematic differences (Random Difference).

Two of the four rater variables in the pretreatment analyses were significantly associated with Level Difference and all were significantly associated with Random Difference. In rater pairs with high Level Difference there was an age difference of 7 years or more and only one rater had some personal therapy or analysis. In rater pairs with high Random Difference, both raters had either seen fewer than 200 cases or one rater had seen over 300 and the other less than 200 cases, one rater had some personal therapy, the other none; one rater completed the IMPS immediately, the other an hour or more following the interview; and finally, there was an age difference of 7 years or more.

As the IMPS items described behaviors indicative of psychopathology and most study patients improved, there was a significant 6-week drop in total variance. There was also a significant drop in intraclass reliability at 6 weeks due to an increase in the percentage of random to total variance. However, none of the eight rater variables was significantly associated with the 6-week Random Difference scores. In contrast, age and experience with personal psychotherapy, the two rater variables associated with Level Difference at the pretreatment period were also significantly associated with the 6-week scores. Amount of psychopathology appears to have a greater influence on Random as compared to Level Difference.

(Received March 19, 1963)

<sup>1</sup> This investigation was supported by Research Grant MYP4663 from the Psychopharmacology Service Center, National Institutes of Health.

An extended report of this study may be obtained without charge from Allen Raskin, Research Department, Department of Psychiatry, D. C. General Hospital, Washington 3, D. C., or for a fee from the American Documentation Institute. Order Document No. 7629 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.



## THE PATTERN OF INTELLECTUAL FUNCTIONING IN PARKINSON PATIENTS

CURTIS HARDYCK<sup>1</sup>

*University of California Medical Center*

AND LEWIS F. PETRINOVICH

*San Francisco State College*

Development of surgical methods for the treatment of Parkinsonism has stimulated interest in the study of intellectual functioning of these patients. The present study compares a factor analysis of the Wechsler Adult Intelligence Scale (WAIS) with results of factor analyses of the WAIS on normal and brain damage subjects. The subjects in the present study were 33 Parkinson patients, representing a cross section of patients with this diagnosis. Mean age of the group is 56, with *SD* of 10.04 and range from 28 to 74. 28 subjects were males and 5 were females.

The present results are compared with results of the following studies: (a) Birren (1952) has reported a factor analysis of the Wechsler-Bellevue (W-B) on 99 subjects, aged 60-74 years. (b) Cohen (1952) has presented a factor analysis of the W-B on 100 subjects with organic brain damage. (c) Cohen (1957) has also presented factor analyses of the WAIS standardization samples. Age groups 45-54 and 60-75+ are compared with the present study.

A principal components factor analysis was done and four factors extracted and subjected to varimax rotation. Results indicate the pattern of loadings for Parkinson patients is similar to both the aged and the brain damage samples, with the greater similarity being with the brain

damage patients. The first factor has high loadings on the Information, Comprehension, Arithmetic, Similarities, and Vocabulary subtests. The aged samples do not obtain high loadings on the Arithmetic subtest as do the Parkinson patients.

The second factor has loadings on Digit Symbol, Picture Completion, Block Design, Picture Arrangement, and Object Assembly. This factor is similar to Birren's aged sample and Cohen's brain damage sample and 60-75+ sample, but differs from Cohen's 45-54 sample.

The third factor loads on Arithmetic and Digit Span and is similar to that reported by Cohen for both the brain damage and 45-54 groups, but differs from Cohen's 60-75+ and Birren's aged sample.

In view of the large differences in the ages of the groups compared, the similarity of the factor structure is quite striking. The Parkinson group is similar both to a group chronologically much older and a much younger sample with a diagnosis of organic brain damage. The factorial structure for the Parkinson patient is least similar to the group most similar in age. This evidence suggests that the changes in intellectual functioning accompanying Parkinsonism are of a fairly drastic nature and possibly will become more extreme with the progress of the disease.

### REFERENCES

- <sup>1</sup> An extended report of this study may be obtained without charge from Curtis Hardyck, Cardiovascular Research Institute, University of California Medical Center, San Francisco 22, California, or for a fee from the American Documentation Institute. Order Document No. 7693 from ADI Auxiliary Publications Project, Photoduplications Service, Library of Congress, Washington 25, D. C. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to Chief, Photoduplication Service, Library of Congress.
- BIRREN, J. E. A factorial analysis of the Wechsler-Bellevue scale given to an elderly population. *J. consult. Psychol.*, 1962, 16, 399-405.
- COHEN, J. Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, 47, 359-365.
- COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, 21, 283-290.

(Received April 29, 1963)

## DREAM REPORT AND SOME PSYCHOLOGICAL CONCOMITANTS<sup>1</sup>

MARTIN S. WALLACH

*School of Medicine, University of North Carolina*

There has been a surprisingly slow accumulation of information relevant to dream experiences, the universality of dreaming, and with respect to psychological correlates of dreaming and reports of dreaming. The present study of dreaming and dream reports partially fills a gap in the current literature by presenting normative data and by suggesting certain psychological concomitants of dream reports.

The sample consists of 332 male college students, 328 of whom are between the ages of 18 and 21. All subjects completed a dream questionnaire; ratings of frequency of dreaming, frequency of dream content recall and pleasantness of dreams; "yes-no" items concerning color dreaming, the self as a character in one's dreams, unreality in dreams, duration of dreams, repetitive dreaming, movement in dreams, and enjoyment of discussing dreams with others; and a description of a dream in as much detail as possible. Also, each subject completed the Minnesota Multiphasic Personality Inventory (MMPI).

Only one subject claims never to dream, 98% of the sample state they sometimes or usually recall dream content and a large majority of the subjects report that about one half or more of their dream life is pleasant. The greater the frequency of dreaming reported the more likely it is that greater content recall will also be reported. In a general way, it can be asserted that a high percentage of the sample proclaim an active and vivid dream life in which they are one

of the characters. Subjects more often than not state their dreams last a long time (71% of subjects), even though "long-time" is not defined, that their dreams go beyond the limits of reality (85% of subjects), and that their dreams are a welcome subject of conversation (77% of subjects). The sample is about equally divided as to admissions of repetitive dreams. Reports of color dreaming, self involved as a dream character, longer duration of dreams, repetitive dreaming, and enjoyment of discussing dreams with others are all statistically significantly related (by chi square) to reports of frequency of dreaming.

Four of the 16 MMPI scales are negatively correlated (although nonstatistically significantly) with reports of dream frequency. Three of these four scales are *L*, *K*, and *R* and lead one to suggest that dream frequency reports are inversely related to defensiveness. Statistically significant positive correlations, for this sample of normals, occur on the *Mf*, *Pt*, *Pa*, *Sc*, and *A* scales ( $r$ 's = .23, .16, .13, .11 and .14, respectively) and lead the writer to suggest that reports of dream frequency are more common among individuals who tend to be introspective, are sensitive to their feelings and inner states, tend to be expressive, and are not overly bound adherents to reality demands. These findings replicate Tart's (1962) conclusion that "subjects classified as Sensitizers report recalling dreaming significantly more frequently than those classified as Repressors."

Dreaming is a usual path to unconscious material, and reports of dreaming represent the product filtered through consciousness. The present study is but a step toward elaboration of the relevance and significance of dreaming which is so often assumed in clinical practice.

### REFERENCE

- TART, C. T. Frequency of dream recall and some personality measures. *J. consult. Psychol.*, 1962, 26, 467-470.

(Received May 13, 1963)

<sup>1</sup>Appreciation is expressed for computational assistance provided by Daniel Brown, and by Ellen Kaplan of the University of North Carolina Computation Center.

An extended report of this study may be obtained without charge from Martin S. Wallach (Department of Psychiatry, North Carolina Memorial Hospital; Chapel Hill, North Carolina) or for a fee from the American Documentation Institute. Order Document No. 7694 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. Remit in advance \$1.75 for microfilm or \$2.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.



## ORIGINALITY IN PROBLEM SOLVING AS A FUNCTION OF ANXIETY AND WITHDRAWAL IN SCHIZOPHRENICS<sup>1</sup>

ALLAN SIDLE

Veterans Administration Hospital, Palo Alto, California

One characteristic of personality disturbance may be a reduced ability to cope with problems in an original manner. To test this possibility, negative relationships were hypothesized between originality in problem solving and two pervasive clinical variables, (a) anxiety and (b) social withdrawal (SW).

Fifty male schizophrenics (mean age, 44; mean hospitalization, 9 years) served as subjects. The subjects were rated by four staff members on anxiety and SW. A paper-pencil inventory (reliability .95), developed by McReynolds and Acker (1962), provided a second measure of anxiety. Two measures of problem solving were used:

1. A "Life Problems Test" (LPT), devised by the author and consisting of 10 situational problems (5 interpersonal in nature, 5 noninterpersonal). The subject was asked to give as many different solutions as he could for each problem. Obtained scores were: (a) *Quantity* (reliability .81), based on the number of solutions given, and (b) *Quality* (reliability .77), based on originality of solutions (in terms of frequency within the sample).

2. Form A of the Howard (1961) Maze Tests (average reliability, as reported by Howard, .80). This consists of several alternate paths to a goal. Each subject completed the maze five times. Score was number of different segments traversed, higher scores reflecting greater originality or novelty.

The means of the interrater reliabilities were .30 for anxiety and .73 for SW. Intercorrelations among the problem solving measures were: (a)

LPT-Quantity versus LPT-Quality .50; (b) LPT-Quantity versus Maze .23; and (c) LPT-Quality versus Maze .26. Only (a) was significant ( $p < .01$ ). The first hypothesis was not supported. The low reliability of the anxiety ratings and the constricted range in the anxiety inventory scores may account for this failure to support the hypothesis. The second hypothesis was supported, the  $r$ 's between SW and LPT-Quantity, LPT-Quality, and Maze being  $-.38$  ( $p < .01$ ),  $-.33$  ( $p < .01$ ), and  $-.30$  ( $p < .025$ ), respectively ( $p$  values based on one-tail tests). The hypothesis is still supported when the effects of intelligence, rated severity of psychosis, and length of hospitalization are taken into account.

It can be further hypothesized that the negative correlation between SW and LPT should be greater for the interpersonal problems than for the noninterpersonal problems. The obtained  $r$ 's were: SW versus LPT-Quantity on interpersonal items .36; SW versus LPT-Quantity on noninterpersonal items .33 (difference between these  $r$ 's NS); SW versus LPT-Quality on interpersonal items .34; and SW versus LPT-Quality on noninterpersonal items .10 (difference between these  $r$ 's is significant at the .05 level on a one-tail test;  $t = 1.76$ ;  $df = 48$ ).

In summary the findings fail to support the first hypothesis concerning anxiety, but they are in accord with the hypothesized negative relationship between SW and originality in problem solving. There is some evidence that socially withdrawn individuals are less able to cope in an original manner with problems of an interpersonal nature than with noninterpersonal problems.

### REFERENCES

- HOWARD, K. I. A test of stimulus-seeking behavior. *Percept. mot. Skills*, 1961, 13, 416.  
MCREYNOLDS, P., & ACKER, MARY. A preliminary report on the development of a technique for assessing anxiety. *Res. Rep. No. 21*, Behavioral Research Laboratory, Palo Alto Veterans Administration Hospital.

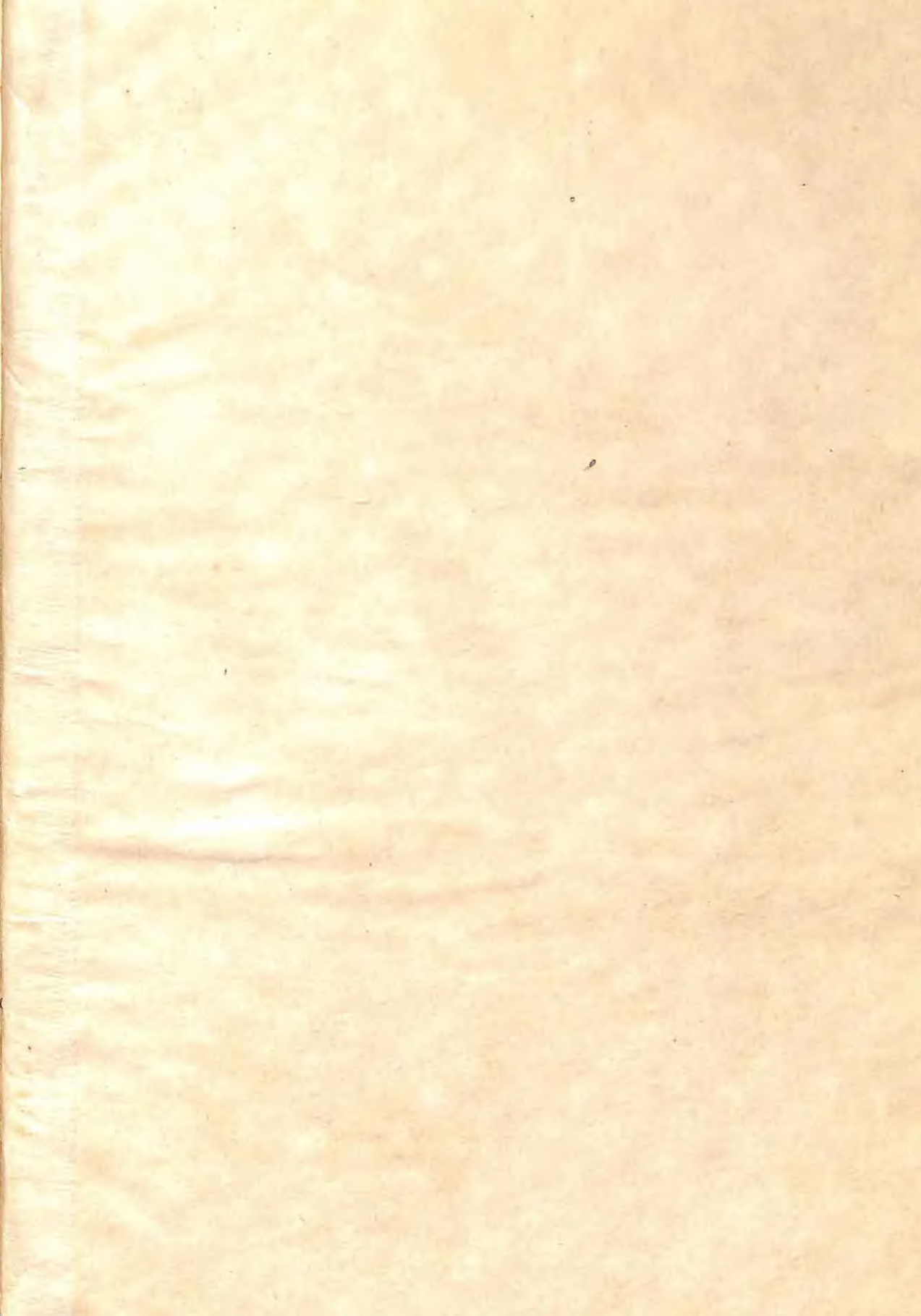
(Received June 25, 1963)

<sup>1</sup> An extended report of this study may be obtained without charge from the Behavioral Research Laboratory, Palo Alto Veterans Administration Hospital, or for a fee from the American Documentation Institute, Order Document No. 7695 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit an advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.











.87 MAR 1972





